

Time for a French revolution

France's scientists take to the streets more readily than most, but are now rightly confronting a neglectful government. They should resist short-term concessions unless these help to secure the long-term health of research.

"The commitment to research must be historic." That's what French president Jacques Chirac pledged in his 2002 election manifesto, calling for science spending to jump from 2.2% of gross domestic product (GDP) to 3% by 2010. His term of office has indeed been historic: historic slashes in science budgets, historic cuts in opportunities for young scientists, and, of late, historic numbers of scientists protesting in the boulevards.

So French scientists fell off their lab stools in disbelief on 6 January when Chirac, in his new-year address, had the gall to reaffirm that research was a national priority, when it was obviously anything but. His speech galvanized a simmering revolt, the Save Research movement, whose petition, launched two days later, has been signed by 66,000 researchers — two-thirds of the entire scientific workforce.

The signatories threatened to resign this week from all management duties, bringing French science to a halt, unless the government redresses a financial crisis at the research agencies, and reinstates 550 full-time research posts that have been turned into short-term contracts (see page 108).

This is far from mere selfishness. The movement is grassroots, emerging independently of trade unions. Its rallying call is that neglect of science is turning away a generation of bright young minds, and that national science policy is adrift. It is right on both counts.

One of its key demands is a thorough overhaul of France's outdated and inefficient science system. The movement has struck a nerve, garnering unprecedented support from the youngest postgraduates to France's star scientists. The momentum for change must not be lost.

Jean-Pierre Raffarin, the prime minister, at first bluntly dismissed the researchers' claims as false, and irritated researchers further by then announcing €1.5 billion (US\$1.9 billion) in tax subsidies to French restaurants. But he is now having to sit up and listen. The

government's calculation that the revolt would peter out has so far proved wrong. For the first time that anyone can remember, research politics is regular headline news and public support is strong.

The administration has been dragged to the negotiating table, and is making concessions. It says it will fork out €294 million of the money it owes research agencies, create 120 full-time posts, boost the value of short-term contracts, and involve researchers in drafting a white paper on science promised by Chirac by the end of the year. Last weekend, Raffarin also pledged an extra €3 billion in research spending, but not until next year. We've heard that one before.

Many in government believe that public spending on research is already sufficient, arguing that it already meets the EU target of 1% of GDP. But this is disingenuous: most goes on France's huge nuclear, space and aeronautics industries. The government has no long-term vision for basic research, and there is little reason to believe it has any intention of developing one. Conservative governments over the past two decades have consistently neglected science. Researchers have good reasons to suspect that the government is seeking to douse the flames of revolt, and then return to business as usual.

The Save Research movement is the best opportunity France has had in a long time to give scientists a strong voice in long-term discussions with government to develop new directions for the country's research, where increased funding is tied to reform of France's bloated and bureaucratic science system. This could culminate in a white paper laying the foundations for a rejuvenation of French science.

To do this, Save Research must remain united. The government's concessions may tempt many scientists to pocket their winnings and return to their labs. But they should hold out for substantial improvements. Better to allow French science to crash, and restart on a sounder footing, than to let it slip slowly into neglectful decay. ■

Enhancing *Nature's* services

Introducing greater opportunity for feedback and improved navigation to relevant literature.

Unsurprisingly, *Nature's* Letters and Articles frequently stimulate responses from the authors' peers. In particular, a reader may submit an attack on the core of a paper. It is then our duty to take it up with authors and referees. If the attack turns out to be well founded, a retraction or correction will follow.

If the criticism is significant but is contested, we publish it in our Brief Communications Arising (BCA) section with a response from the authors. Until now, space limitations have restricted our ability to publish all but the most significant of such debates. But from next week's issue we shall be increasing the number of such BCAs by publishing them online only, drawing attention to them by their titles and authors in the print edition's Brief Communications section and our table of contents. The BCAs will be citable by their Digital Object Identifiers, like all *Nature* content, and forward-linked from the original paper. This expansion will also allow us occasionally to publish comments that are more positively stimulating.

Such feedback needs no prompting. In contrast, in another

enhancement, we wish to solicit feedback from authors of our papers in order to improve the service that they receive. Thus, from 1 January this year, all authors are being invited to complete an online feedback form about aspects of editorial services, from refereeing quality to efficiency and transparency in handling.

And there are two improvements in navigation for readers. One relates to our research papers. Many carry additional information for specialists in their online Supplementary Information, which has become more extensive in recent years. Readers will find that the presentation of this has been redesigned for quicker and easier access.

And finally (for now), we have introduced new internal search (or 'semantic matching') technology that allows us to scan the complete online content of *Nature* journals and automatically generate updated web links to related articles on request via a click on the left-hand margin of online *Nature* content.

As always, *Nature* welcomes readers' comments about these enhancements, and suggestions for others. ■





Striking out
Academics and students take action across Europe
p108



Guided weapons
US watchdog helps biologists police 'dual use' research
p109



Tag team
Conservationists split over scheme to save dolphin
p111



On thin ice
Russian researchers rescued from sinking Arctic research base
p112

Transgenic planting approved despite scepticism of UK public

Jim Giles, London

After years of careful deliberation, the British government this week took a crucial step towards acceptance of genetically modified crops.

On 9 March, government ministers announced limited approval for commercial planting of a maize (corn) variety engineered to be resistant to a specific herbicide — the first such approval in Britain.

Although the approval lasts only until 2006, the decision indicates that prime minister Tony Blair's government is willing to support transgenic technology in the face of widespread public opposition. In Britain, opposition to agricultural biotechnology has been early and strident. Both supporters and enemies believe this week's decision will influence debates outside Britain about transgenic crops (see below).

Environment minister Margaret Beckett announced that farmers can grow the maize if they follow management guidelines used in farm-scale studies of the crop's environmental impact. These studies found that the maize could have a beneficial effect on biodiversity, provided the timing and nature of herbicide spraying are carefully controlled (see *Nature* 425, 751; 2003).

Blair's government has generally backed the technology since coming into office in 1997, but it has been cautious about allowing commercial cultivation. Opinion polls show that the public is concerned about the impact of the crops on human health and the environment, and environmental groups have campaigned against the technology.

The case for the crops was boosted by a scientific review, released last July, which found no reason to rule out carefully managed cultivation of the plants. The review was discussed at a cabinet meeting last month. Leaked minutes of the meeting state that ministers acknowledged public opposition, but thought that it "might eventually be worn down by solid, authoritative scientific argument".

Despite this week's announcement, the company marketing the maize used in the farm-scale study — Bayer CropScience, based in Monheim am Rhein, Germany — is



Seeds of change: British farmers could start planting genetically modified maize in 2005.

unlikely to sell many seeds in Britain in the near future. The maize has had European Union approval since 1997, but the terms of the approval will have to be modified to take

the herbicide restrictions into account before sales can start. Officials in Beckett's department say this will take several months, although they predict that the necessary rules will be in place for farmers to plant transgenic maize in spring 2005. Farmers will, however, be keeping a nervous eye on the first plantings, as environmental activists have damaged several research plots.

Farmers will also be wary of planting genetically modified varieties before the government has clarified rules governing how they should be kept separate from nearby conventional crops. In addition, European Union approval for the maize will have to be renewed if cultivation is to extend beyond 2006.

Renewing the application may not be a formality. The farm-scale trials compared transgenic maize with conventional varieties that had been treated predominantly with triazine herbicides, powerful weed-killers that are likely to be banned in Europe by 2006. When compared with fields treated with less intense herbicides, the biodiversity benefits associated with transgenic maize decrease by around a third, according to a study led by Joe Perry, a statistician at Rothamsted Research, an agricultural research firm based north of London (see J. N. Perry *et al.* *Nature* doi:10.1038/nature02374; 2004). ■

Californian county bans transgenic crops

A rural county in northern California voted last week to become the first county in the United States to ban the planting of genetically modified crops.

Voters in Mendocino County, 130 kilometres north of San Francisco, approved the ballot initiative on 2 March by 56% to 44%, despite a US\$700,000 campaign against the ban by supporters of the technology.

Similar ballot initiatives are planned in nearby Humboldt and Sonoma, and activists in half a dozen other counties in California are considering similar actions.

A state-wide ballot initiative is also being pursued in North Dakota against the planting of genetically modified wheat. The initiative is being led by a group of wheat farmers who say the crop

could contaminate other strains and sabotage their export markets in countries such as Japan.

The grass-roots drive in Mendocino County was led by Els Cooperrider, a retired cancer researcher, and her husband Allen, a former zoologist, who run an organic brewery there. No transgenic plants are currently grown in the county, where grapes for wine are the main crop.

Peter Bradford, a cattleman and president of the Mendocino County Farm Bureau, says the vote was in part driven by "a fear of science and big corporations". Grape growers will probably use the ban as a marketing tool, he adds.

Agricultural corporations are considering a legal challenge to the county ban, or pursuit of state wide legislation to overturn it.

Rex Dalton, San Diego

Wave of protest strikes Europe's universities

Laura Nelson

Discontent is sweeping through European universities this winter, as academics protest against stagnant salaries, dwindling career prospects and increasing demands made on them by their employers.

In France, Italy and the United Kingdom, the protests have drawn tens of thousands of scientists into a spree of militancy not seen for decades. Many doubt that their protests will have much impact on their politician paymasters. But, disillusioned by years of what they see as disrespect for their chosen vocation, most are joining in the protests — if only to make the point that they are unhappy with the status quo.

In Britain, a proposed fee structure that will end the tradition of free tuition for undergraduates has combined with an unwelcome new pay structure for academics to trigger campus unrest. The Association of University Teachers (AUT), the main academic teachers' union, called its 48,000 members out on one-day strikes during 23–27 February, and says that 90% of them took action.

On a picket line outside University College London on 25 February, student supporters of the action were in boisterous mood, but the handful of dutiful scientists present found little to smile about. "We're all a bit disillusioned," grumbled John Pollard, an electrical engineer, who said that he has never been on strike before. Other strikers said that they were frustrated at what they



All out: union leader Sally Hunt addresses pickets at Cardiff University during a week of action.

view as the government's general neglect of higher education.

The AUT says that its members are putting up with the low pay that accompanies academic work, while their increasingly regimented working environment has come to resemble that of better-paid peers in the private sector. "We want respect for people who have been delivering against every benchmark they have been set for the past decade," says Sally Hunt, the union's general secretary.

Other scientists mutter that the protests won't help. "Of course academics aren't paid enough," says one anonymous neuroscientist at the University of Oxford. "But striking won't do any good. Unlike a bus driver,

nobody cares when a scientist goes on strike."

The AUT has sought to increase its leverage by asking its members to boycott student assessments, including the acceptance of PhD theses, from 1 March. But scientists are not keen to take this step. "It is unlikely that we will boycott assessments," says Pollard. "We are too committed to our students."

French academics are taking a more aggressive approach to their grievances. Half of the country's scientific administrators were threatening to resign from their management duties this week in protest at low research funding levels and job cuts (see left). "This is the first time the protests have been on such a large scale," says Alain Trautmann, a cell biologist at the Cochin Institute in Paris.

In Italy, most young academic scientists took part in two one-day strikes in February and March to protest at a draft law, released on 16 January, that restructures professorships, sharply increases minimum teaching hours and assigns control of university posts to the government. Giovanna Grimaldi, a geneticist at the Institute of Genetics and Biophysics in Naples, says that the proposed change "breaks the autonomy of universities", adding that the most worrying aspect of the plan is that it would "control people" and take away their independence.

All this has led to an explosion of frustration, says Grimaldi, which she thinks is damaging the attractiveness of a scientific career. "Now young scientists go away to the United States, and they don't come back," she says.

This brain drain of talent from European science — driven in large part by greater salaries and career opportunities in the United States — is a theme commonly aired by protesters in each country. Peter Cotgreave, director of the pressure group Save British Science, says that a widespread sense that "the grass is greener in the United States" is helping to keep European researchers feeling blue. ■

French scientists prepare for mass resignation

French scientists were meeting in Paris this week to decide whether to go through with their threat to resign all management duties. The ultimatum hinged on the government taking immediate and significant steps to boost French research. But as *Nature* went to press, the movement's leaders remained adamant that many researchers would quit.

The mass resignation would be the culmination of a bitter winter of protests by researchers in government laboratories and universities. They say that President Jacques Chirac and Prime Minister Jean-Pierre Raffarin have failed to keep funding promises, or to consult with them about a strategy paper that is being prepared on the future of French research.

The resignation pledge was instigated by a grass-roots body called Save Research (see *Nature* 427, 276; 2004), which has so far attracted some 66,000 signatures. If implemented, it could quickly paralyse the laboratory system: lab heads, for example, are legally responsible for security, so their departure could mean that labs will be unable to open their doors. It may also become impossible for staff to

order supplies or get permission to travel.

In the run up to this week's meeting, the government announced concessions to head off the revolt, including an offer of a multi-year funding plan for research, but these have so far been rejected by the protesters (see page 105). On 5 March, Raffarin and science minister Claudie Haigneré discussed the crisis with Étienne-Émile Baulieu and Édouard Brézin, the president and vice-president, respectively, of the French Academy of Sciences.

Baulieu and Brézin have offered to mediate between the government and the protesters. They proposed the creation of an independent high-level committee, including members of Save Research, to ensure scientific participation in the strategy paper, plans for which were announced at the start of the year by Chirac. Raffarin accepted the idea, says Brézin.

The promise of a funding plan for science would usually hearten researchers, Brézin notes, but not in the current climate. "There is a crisis of confidence in the government," he says.

Declan Butler

♦ <http://recherche-en-danger.apinc.org>

Terror watchdog set up for 'dual use' biology

Erika Check, Washington

The US government has created a high-level advisory body that will suggest guidelines for scientists whose work might be used by bioterrorists.

Since the terrorist attacks of 11 September 2001, some biologists have worried that the federal government might respond to concerns by imposing a clamp-down on their work. So the latest plan has been cautiously welcomed as an indication that the life-sciences community will be allowed to police itself.

On 4 March, Tommy Thompson, the health secretary, announced the creation of a National Science Advisory Board for Biosecurity. The board will be appointed by Thompson's office and is expected to be up and running by the autumn.

As well as giving advice to government departments about how to approve and manage 'dual-use' projects, which have the potential to be used in bioweapons, the board is expected to advise publishers on policies for handling sensitive papers, and researchers on what they can talk about at open scientific meetings.

None of the recommendations will be mandatory, government officials say — but agencies may choose to make adherence to the guidelines mandatory for their grantees.



Helping hand: Elias Zerhouni, Anthony Fauci and Tommy Thompson launch a National Science Advisory Board for Biosecurity that will guide researchers.

Thompson and John Marburger, president Bush's science adviser, said that the board is being created partly in response to a report by the US National Academy of Sciences. Marburger added that the administration has been concerned by a series of recent research reports with dual-use potential, such as a paper about the creation of a synthetic polio virus (J. Cello *et al. Science* 297, 1016–1018; 2002).

Elias Zerhouni, director of the National Institutes of Health, whose office will house the board's secretariat, said that its workings will be modelled largely on those of the Recombinant DNA Advisory Committee (RAC), which was created in 1974 to imple-

ment rules for conducting studies using recombinant DNA.

Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, said the RAC has worked because it created a "culture of responsibility" that has been adopted by researchers in universities and industry. Its rules are generally implemented by local institutional biosafety committees, which will also implement the new guidelines, officials said.

Biologists' groups have cautiously welcomed the plan. "We haven't seen a complete picture, but it seems to sponsor open research and involvement of scientists," says Janet Shoemaker, director for public and scientific affairs at the American Society for Microbiology.

But some security experts object to the fact that the advisory board will not have the power to make compliance with its rules mandatory. "I find it very curious that they accept the need for mandatory oversight of access to pathogens but reject the idea of mandatory oversight of consequential work with those pathogens," said Elisa Harris, pathogen projects coordinator at the Center for International and Security Studies at Maryland in College Park. "It's the work that has the greatest potential to cause destruction or harm on a catastrophic scale." ■

Nigerian states disrupt campaign to eradicate polio

Declan Butler, Paris

A boycott by two Nigerian states of an immunization campaign by the World Health Organization-led Global Polio Eradication Initiative could jeopardize its goal of eradicating the disease by the end of this year.

Kano and Zamfara states refused to take part in a February campaign by the World Health Organization (WHO) in west and central Africa that aimed to immunize 63 million children. Muslim clerics there have claimed that the vaccines are contaminated with contraceptives, carry HIV and are a Western plot against their populations.

Some community leaders are also saying that the vaccine causes AIDS, having picked up on an international debate over the now-refuted claim that it was the origin of the AIDS virus (see *Nature* 404, 9; 2000). "If you search for 'polio vaccine' on the Internet you find a lot of stuff — not all of it good," says Oliver Rosenbauer, a WHO spokesman.

Such claims led several states in northern

Nigeria to boycott a WHO campaign in August. Since then the federal government has backed immunization, and in October, the Organization of the Islamic Conference, which represents 56 Muslim nations, gave its support to the polio-eradication initiative.

But two uncooperative states is still too many, warns Rosenbauer, who says that eradication must take place in all territories. The virus must be eliminated wherever it remains, he warns, as a single case could spawn an outbreak.

Indeed, last summer's boycott by several of Nigeria's northern states is seen as responsible for the country having the largest number of new polio cases — 347 from the start of 2003 to February 2004, nearly half of all new cases worldwide. It is also blamed for 20 new polio cases in recent months in 8 previously polio-free African countries: Benin, Burkina Faso, Cameroon, Central African Republic, Chad, Côte d'Ivoire, Ghana and Togo.

Nigeria has since set up joint team of scientists and community leaders to look at the vaccine's safety. It is expected to release its report officially later this month. WHO officials are optimistic that this will be in time for the two states to consider their position ahead of the next round of immunization in late March.

"All the indications are that the tests have proved negative for any harmful or suspicious agents, and in particular, none of those allegedly capable of promoting birth control were detected," says Idris Mohammed, chairman of Nigeria's National Programme on Immunization.

But even if all Nigerian states comply, the publicity surrounding the rumours has sown seeds of doubt about polio vaccination among some community leaders. To what extent this affects the latest campaign will only become known later this month, when an analysis of data coverage is completed. ■

♦ www.polioeradication.org

Biotechnologists seek to bridge South Asian divide

K. S. Jayaraman, Hyderabad

Biotechnology collaborations have been agreed between India and Pakistan, ending more than five decades of scientific impasse between the two nations.

Last week, a Pakistani delegation to the BioAsia 2004 meeting in Hyderabad signed five agreements — three with Indian biotechnology companies and two with the All India Biotech Association (AIBA). The agreements are said to be the first involving a high-technology area to have been struck between the rivals.

“I will go back to my country extremely pleased,” says Anwar Nasim, chairman of Pakistan’s National Commission on Biotechnology, who led the delegation.

“We intend to collaborate with Indian companies in the areas of industrial products, vaccines, diagnostic kits and transgenic crops,” Nasim told *Nature*. Under the agreements, the AIBA will help Pakistan to establish its own biotechnology association and will provide a list of Indian technologies that are available for licensing in Pakistan.

B. S. Bajaj, a senior AIBA official, says that in the long term the partnership should benefit both nations. “Indian technology could help Pakistan to bring down the price of its drugs, which are six to seven times more costly than in India,” he says. “And Indian biotechnology companies will benefit from opening up a new market in their backyard.”

The agreements are general in nature, however, and “we will have to see how they progress”, cautions the AIBA’s Ashok Sadim Khan. Nasim agrees that “complex issues will have to be dealt with” before the agreements are implemented. But he adds: “All I can say is that we are making a start, and are confident that many avenues will open up for collaboration.”

Nasim says that prospects for improved relations between Indian and Pakistani scientists are looking brighter. “Our academy has already received an invitation from the Indian National Science Academy (INSA) and we are considering it,” he adds. INSA secretary S. K. Sahni told *Nature* that, because things move slowly at government-to-government level, the INSA had decided to offer to launch discussions with the Pakistan Academy of Sciences on such common interests as agriculture and malaria. ■



Problems of the poor set to face cost-benefit treatment

Jim Giles, London

Björn Lomborg was vilified by the green movement when he published *The Skeptical Environmentalist*, a 2001 book that questioned the validity of several widely held beliefs about the state of the planet. A furious activist went so far as to thrust a pie in his face, on at least one occasion. But his next big venture could see him rouse the ire of an equally passionate group: the international aid movement.

As part of Lomborg’s latest scheme, nine eminent economists will gather in Copenhagen on 24–28 May to rate solutions to the developing world’s ten largest problems, from financial instability to communicable diseases (see above). The panel intends to use cost-benefit analyses to evaluate between three and five rival approaches to each problem. The result, says Lomborg, will rank the most cost-effective ways of doing good.

It could also be the recipe for a major row. International charities say the scheme ignores established targets, such as those developed by the United Nations. Some economists say that Lomborg is overstating the usefulness of cost-benefit analysis — a technique that has long been used by conservative economists to support arguments against everything from clean-air rules to development aid for the poor. “To believe the problems of the world can be solved like this is absurd,” says Eric Neumayer, an expert in environment and development at the London School of Economics. ■

The economists planning to take part in the assessment mostly reside in elite universities in Europe and the United States, and four of them are Nobel laureates. They will each assign numbers to potential solutions, of which there are around 30; the numbers will represent the size of the benefits that come with the approach, such as fewer cases of disease, minus the costs of implementing it.

Cost-benefit analyses have become more common in aid planning over the past decade, but remain controversial nevertheless. Economists have to assign values to variables in the analyses, such as human life, and the numbers they use can vary widely. Critics such as Neumayer say that for some problems, such as climate change, the variability is far too great for the results to be meaningful.

Others point out that the quantitative approach, if it were adopted by aid agencies, would inevitably push effort away from the toughest problem areas — such as war-torn southern Sudan. “There is little chance of effective use of money there,” says Roger Riddell, international director of Christian Aid, a London-based charity. “But it should not be abandoned.”

Lomborg acknowledges that some countries could lose out, but argues that we already prioritize aid funding and that this is simply a better way of doing it. “There’s never an easy answer,” he says. “But if there are places where we can do relatively little then we should consider achieving more elsewhere.” ■

♦ www.copenhagenconsensus.com

Air force clips the wings of UK wind power

Laura Nelson, London

Britain's bold plans to generate more electricity from wind energy may be shot down by the Royal Air Force (RAF), which says that wind farms confuse its air-defence radar system.

But wind-energy advocates are pressing the air force to ease up on rules that currently prohibit wind turbines from being located within a radius of 74 kilometres of any of its 13 air-defence radar stations.

Earlier this month, the Royal Society, Britain's science academy, intervened in the dispute. It wrote to both the Department of Trade and Industry (DTI) and the Ministry of Defence (MoD) to point out that the RAF's stipulation could derail government plans to generate 10% of electrical power from renewable sources by 2010.

David Wallace, the society's vice-president, wrote in the letter that nearly half of the wind farms proposed so far in Britain have been successfully opposed by the MoD because of their proximity to air-defence stations. Wallace said that the restrictions were a "significant obstacle on the prospects for developing the UK wind industry".

On 1 March, the society gave written evidence to an inquiry by the House of Lords Science and Technology Committee on renewable energy. The society said that it was concerned about the number of applications for wind farms that had been turned down.

The MoD says that the 30- to 40-metre-long blades used on typical wind turbines can mask radar signals from low-flying aircraft or produce false alarms by reflecting radar



Outmanoeuvred: radar stations struggle to tell the difference between low-flying jets and wind turbines.

beams. The wind-farm restriction is based on two RAF studies carried out in 1994 and 1997, the contents of which remain classified.

Other European countries do not impose such restrictions. Germany, for example, demands a distance of only 5 km between wind farms and radar stations. But MoD officials say that there are more low-flying aircraft in the United Kingdom.

A working group of representatives from the MoD, the DTI, the Civil Aviation Authority and the British Wind Energy Association

was set up two years ago to address the problem. But it has so far failed to agree on whether the MoD rule is technically justified.

Instead, the group has discussed ways to avoid the problem, such as using radar systems that can distinguish between wind turbines and aircraft. The MoD may implement the more sophisticated radar in 2008, but the system is expensive and the ministry does not believe it will completely solve the problem. The Royal Society is urging government ministers to step in and resolve the dispute. ■

Scheme to track rare dolphins hits troubled waters

David Cyranoski, Auckland

A plan to help save one of the smallest and rarest species of dolphin has divided conservation biologists.

New Zealand's conservation department wants to track between 10 and 50 Maui's dolphins (*Cephalorhynchus hectori maui*) by attaching a satellite tracking device to their fins. But critics are concerned that the plan could do more harm than good.

Only about 150 of the 1.7-metre long dolphins remain, and government officials say that the tags could be used to direct fishing boats away from the mammals.

In the past three years, seven Maui's have been washed up on New Zealand's shores, probably as a result of becoming tangled in fishing nets, says Rob McCallum, a senior scientist at the conservation department. "In a population of 150 dolphins, that's a lot," he says.

But the 50-gram transmitter tags, which

are fixed to a fin by two nylon pins, could upset and injure the dolphins — or even kill them, say critics. Maui's are much smaller than most of the other mammals on which the technology has been tested, they note.

In a letter to McCallum last November, Nick Gales, a biologist at the World Conservation Union, called for a "thorough risk assessment" before the technique was



Hector's dolphins (above) will be used to test technology to track the rarer Maui's dolphin.

used on the Maui's. "The level of concern is particularly high for the smaller species, given that the potential impact of the tag is related in part to the size of the animal," he wrote.

And in a report this January, Ian West and Simon Childerhouse, marine-mammal specialists at the conservation department, questioned how useful the tagging data would prove to be. They recommended that non-invasive methods, such as aerial sightings, be used instead.

But on 5 March, the conservation department began a three-month trial to test both the safety and usefulness of the tagging system. Researchers have tagged three Hector's dolphins, which are close relatives of the Maui's and are roughly the same size.

McCallum, who is managing the trial, says that satellite tracking has to be tested, even if it carries some risks. "We are not prepared to just watch as this species slides into extinction," he says. ■



Life-saver: a Russian helicopter flies in to rescue a team of researchers from a sinking ice floe.

Russian researchers on thin ice rescued as Arctic base cracks up

Munich In a dramatic helicopter operation on 6 March, 12 Russian researchers were rescued from a sinking ice floe 500 km south of the North Pole. The team had been operating a free-drifting manned polar station.

The ice-pack broke into two sections on Wednesday. The part containing most of the camp then sank without trace. In temperatures below -20°C , and with supplies for only three more days, the team waited for help in a small building in the intact part of the station before being rescued and taken to St Petersburg. None of the researchers was injured.

According to the initial schedule, the station — North Pole 32 — was to be abandoned on 28 March, and the team collected by a Russian nuclear ice-breaker. Despite the drama, a new free-drifting station, North Pole 33, is expected to be launched in September.

Authors retract link between MMR vaccine and autism

London A speculative link between the combined measles, mumps and rubella (MMR) vaccine and autism has been formally withdrawn from a controversial paper by most of the original authors.

Use of the vaccine dropped in Britain after lead author Andrew Wakefield, a gastroenterologist then at the Royal Free Hospital in London, suggested a link between the vaccine and autism (A. Wakefield *et al. Lancet* 351, 637–641; 1998).

Wakefield was accused last month of failing to acknowledge at the time of publication that he had received funding for a study to investigate whether parents whose children had allegedly been harmed by the MMR vaccine could sue the manufacturers.

Many of the co-authors had already distanced themselves from the paper but, in a statement published last week, 10 of the 12 formally retracted the part of the paper that

linked the vaccine with autism (S. H. Murch *et al. Lancet* 363, 750; 2004).

Beijing and San Diego team up for research

San Diego Beijing University and the University of California, San Diego, have formed an alliance to advance medical genetic research at a US\$14-million institute in China.

The Institute of Molecular Medicine China will initially focus on cardiovascular disease studies, later expanding into metabolic and neurological fields. Both basic research and clinical studies are planned.

The director of the institute will be Rui-Ping Xiao, a Chinese physician and cell-physiology researcher currently at the US National Institute on Aging in Maryland. “Our dream is to build an institute to fight China’s number one health problem: cardiovascular disease,” says Xiao.

Britain accelerates push towards linear collider

London Britain took a step towards becoming a major player in an important planned particle-physics experiment: the international linear collider.

The Particle Physics and Astronomy

Research Council last week revealed plans for the Cockcroft Institute, a national centre for accelerator science, to be established at Daresbury Laboratory in Cheshire as part of an eight-year, £21-million (US\$39-million) scheme. A second centre will be split between the University of Oxford and Royal Holloway, part of the University of London.

The project should help the country's scientists participate in the linear collider, a multibillion-dollar study of particles produced by high-energy collisions between electrons and positrons. A final decision on funding for the collider may be made as soon as 2007.

The council has also boosted the hopes of physicists trying to build a neutrino factory (see *Nature* 418, 117; 2002), allocating £3 million to develop technology for it.

Ecstasy earns medical study after rave reports

Washington A long-delayed US clinical trial of the clubbers' drug ecstasy has cleared its final regulatory hurdle and looks set to begin this spring.

The trial is sponsored by a non-profit organization called the Multidisciplinary Association for Psychedelic Studies (MAPS), based in Sarasota, Florida. It will involve patients who suffer from post-traumatic

stress disorder, who will undergo sessions of psychotherapy with and without 3,4-methylenedioxymethamphetamine, the active ingredient in ecstasy. The study builds on case reports from patients who say they were able to overcome flashbacks, nightmares and other post-trauma symptoms while in therapy aided by ecstasy.

The study is the first in the United States to test a therapeutic use for ecstasy since the drug was criminalized there in 1985. The study was first approved by the US Food and Drug Administration in 2001, before being surrounded in dispute when researchers at Johns Hopkins University suggested that a single dose of ecstasy could kill a monkey. Their paper was retracted in September 2003, clearing the way for US government agencies to approve the study. MAPS investigators received a licence from the Drug Enforcement Administration at the end of February.

Philanthropist funds space and nanotechnology labs

Washington The Kavli Foundation has announced the opening of seven research institutes at universities in the United States and Europe, with \$100 million allocated for nanotechnology, cosmology and neuroscience research.

Norwegian-born physicist and philanthropist Fred Kavli, who formed the foundation, believes that these fields "provide the greatest opportunity for major scientific breakthroughs and will have long-range benefits for humanity". Institutes at Cornell University, the California Institute of Technology and Delft University of Technology in the Netherlands will focus on nanotechnology, while facilities at Columbia University, Yale University and the University of California, San Diego, will focus on neuroscience. A cosmological institute at the University of Chicago will join two existing theoretical physics centres at the University of California at Santa Barbara and Stanford University.



Finding funds: Fred Kavli will spend some \$100 million on US and European research.

A rising tide

The ice covering Greenland holds enough water to raise the oceans seven metres — and it's starting to melt. How far will it go? Quirin Schiermeier wades into the evidence.

Greenland was not always covered with ice. About 60 million years ago, when Earth's climate was much warmer than today, the world's largest island was a grassy arctic tundra across which the ancestors of modern horses and other mammals migrated.

Since that time, millions of cubic kilometres of ice have accumulated on the island. No longer home to migrating beasts, Greenland now serves as a reservoir for immense quantities of water. And it would be best for the water to stay there. If Greenland's ice sheet were to melt, the oceans would rise by seven metres, enough to cover large parts of low-lying countries such as Bangladesh and the Netherlands. New York cabbies would have to buy motor boats. Most of southern Florida, from Key West to the Everglades, would sink beneath the waves.

This scenario is not just fantasy. It has become clear in recent years that climate warming is beginning to trigger rapid and substantial changes in the polar ice caps. In Greenland, these changes may be irreversible. New research suggests that if Greenland's ice sheet melts, it will probably not form again, even if temperatures and heat-trapping greenhouse gases returned to pre-industrial levels.

Although the Greenland ice sheet is much smaller than its counterpart in Antarctica, where some ice loss has also been noted, it now seems more likely to contribute to a rise in sea level. For one thing, floating sea ice has been on the decline in the north in recent years. With less ice around to reflect sunlight, the oceans are absorbing more energy than they were before, leading to further warming. In contrast, southern sea ice has been relatively stable. The lower average temperature in Antarctica means that ice melting off glaciers refreezes before it can reach the sea. But Greenland's warmer summers allow glacial melt to pour into the North Atlantic, further accelerating the loss of ice.

The bulk of Greenland's ice, more than



All change: melting ice carves channels through Greenland's ice sheet, increasing the movement of ice towards the sea.



3,000 metres thick in the centre, is a remnant of the last ice age, which peaked some 20,000 years ago. It still exists because temperatures in the Holocene — the period from the end of the last glacial epoch to the present — have never risen beyond a point where the ice lost to melting, evaporation and the march of glaciers to the sea has on average exceeded the accumulation of new ice through snowfall. The ice sheet has been in balance.

Slippery slope

But there is growing evidence that the system is about to lose its equilibrium. Philippe Huybrechts, a glaciologist and ice-sheet modeller at the Free University in Brussels, has modelled the behaviour of Greenland's ice sheet under various climate scenarios. According to projections, Greenland's mean annual temperature could increase by 8 °C

by 2100. Huybrechts found that this would mean that by the end of the millennium, the ice sheet would shrink to a small glaciated area far inland and the sea level would rise by six metres. This is a worst-case scenario, says Huybrechts, who reported his findings in the third assessment report of the Intergovernmental Panel on Climate Change (IPCC), but it is not outside the realms of possibility. Substantial shrinking can be expected even for more conservative temperature rises, Huybrechts says¹.

Because a certain amount of global warming now seems inevitable in response to greenhouse gases already released through human activity, an important question is whether Greenland's ice sheet, which has already begun to disintegrate, will reach a point beyond which it cannot recover.

One body of work suggests that it may. Because the ice is so thick, much of Greenland's heartland lies at a very high altitude. This, in turn, ensures sufficient annual snowfall to build up new ice as fast as the sheet degrades near the coast. But modelling by Jonathan Gregory, a climate researcher at the Centre for Global Atmospheric Modelling in Reading, UK, suggests that as ice is lost, portions of the surface of Greenland's interior will end up at lower elevations where the air is warmer. Less snowfall and more rain would cause the ice to disappear at a



CORBIS
Hotting up: the temperature on Greenland looks set to rise by some 3 °C by the end of this century.

faster rate than it was being replaced, leading in turn to further drops in elevation.

In one simulation carried out at the Hadley Centre for Climate Prediction and Research in Bracknell, UK, Gregory and his team modelled the topography of Greenland after completely removing the ice sheet. They then simulated cool pre-industrial climate conditions — which would easily maintain the existing ice sheet — to see whether the ice would regenerate. It did not². The ice sheet, they concluded, creates its own local climate. It depends on itself to exist.

“The ice is only there because it has always been there since the end of the last ice age,” agrees Huybrechts. “If it had melted at some point in the past 15,000 years it would not have come back.”

Losing out

Meanwhile, satellite and aircraft surveys as well as measurements by scientists on the ground are painting a picture of an ice sheet that is rapidly losing mass. Although data for some regions are still poor, it is clear that most of the loss is taking place near the coasts³.

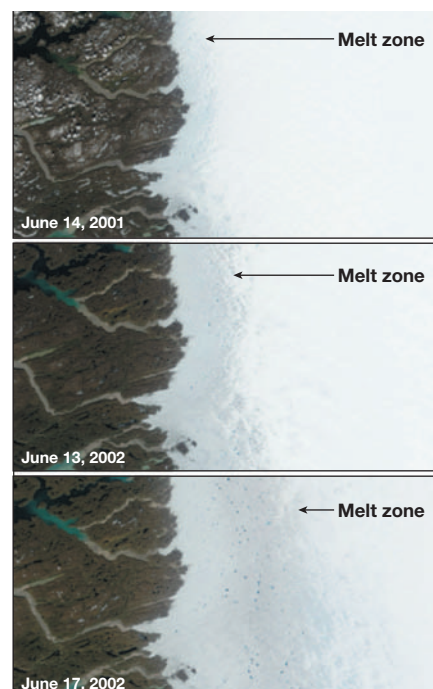
Bill Krabill, a project scientist at NASA's Goddard Space Flight Center in Wallops Island, Virginia, coordinated aerial surveys of the ice sheet's elevation from 1994 to 1999. Aircraft flying at a steady altitude aimed a laser at the surface and calculated the height of the ice based on the time it took for the laser beam to bounce back. As ice is lost from the edges first, Greenland's ice sheet, a shallow dome, is effectively getting steeper. Krabill's team found that over five years, ice around the edges disappeared faster than before⁴. Subsequent measurements have continued to support that conclusion, Krabill says.

One reason the pace may be accelerating is that the melted ice does more than just run

into the ocean. It also seeps down to the ice–bedrock interface, where it acts as a lubricant, quickening the pace of ice movement towards the sea. As the glaciers reach the margins of the island, great icebergs split off into the water.

“Along the coasts, all the glaciers but one are thinning like mad, and they're also flowing faster than they ought to,” says Eric Rignot, a glaciologist and principal scientist at the Jet Propulsion Laboratory in Pasadena, California.

Scientists are less certain about what is going on in the central area of the ice sheet. Is



Signs of erosion: these yearly photographs clearly show the increased melting of coastal ice on Greenland.

it getting thicker or thinner? The key will be to find out whether the changes near the edges are just normal fluctuations or the beginning of a long-term trend, says Rignot. “Changes initiated in coastal regions will propagate inland very quickly,” he believes.

Krabill estimates the rate of ice loss from Greenland to be about 50 cubic kilometres per year, or roughly 1/50,000 of its total volume. That should produce enough water to raise the global sea level by an average of 0.13 millimetres per year⁵. This is not yet an acute threat, says Krabill, but he notes that his estimation of net mass loss is conservative. “It could easily be twice that much,” he says.

Heated debate

A number of factors complicate forecasts about the ice sheet's future. Computer models suggest that if Greenland warms up by 3 °C in the next 100 years — which is well within the projections of the IPCC — the ice sheet cannot survive⁶. But these models depend on assumptions about the likely rate of ice retreat that are not guaranteed to be correct, in part because the flow dynamics of glaciers are not yet fully understood. In Antarctica, for example, ice streams seem to come and go, and the movement of glaciers has been seen to speed up and slow down very quickly. But the cause of such shifts is not always clear.

Further complicating the models are the contradictory effects a warmer climate may have on polar ice sheets. As temperatures rise, evaporation from the surrounding oceans will increase, sending more moisture inland. This will increase snowfall over high-elevation accumulation zones. On the other hand, the simultaneous increase in summer rains could accelerate the melting. So the net effect of increased precipitation is hard to predict.

One final uncertainty is the fate of the Gulf Stream, which delivers warm tropical water to the North Atlantic and keeps Europe much warmer than it would otherwise be. Some scientists believe that the Gulf Stream is weakening in response to the massive influx of fresh water from shrinking glaciers. If it became weak enough, they say, Greenland would chill down again and its ice sheet might stabilize.

Fresh water is stored as ice all over the world, not just in ice sheets but also on inland mountain peaks. In theory, if all of it melted, the oceans would rise an astounding 80 metres. Fortunately, climate experts agree that this is not on the cards. But the uncertainty over Greenland is just one more reason not to put off that vacation to Florida. ■

Quirin Schiermeier is *Nature's* German correspondent.

1. Huybrechts, P. & de Wolde, J. *J. Clim.* **12**, 2169–2188 (1999).
2. Toniazio, T., Gregory, J. M. & Huybrechts, P. *J. Clim.* **17**, 21–33 (2004).
3. Rignot, E. & Thomas, R. H. *Science* **297**, 1502–1506 (2002).
4. Zwally, H. J. *et al. Science* **297**, 218–222 (2002).
5. Krabill, W. *et al. Science* **289**, 428–430 (2000).
6. Huybrechts, P., Letreguilly, A. & Reeh, N. *Glob. Planet. Change* **3**, 399–412 (1991).



Growing old gracefully

Across the industrialized world, birth rates are falling and people are living longer. This will require a new focus on research to promote healthy ageing, rather than simply treating the diseases of old age. Alison Abbott reports.

Thomas More's *Utopia*, published in 1516, describes an egalitarian paradise free from the poverty and wanton greed of contemporary sixteenth-century European society. But its fictional citizens are still haunted by old age — "which as it carries many diseases along with it, so it is a disease of itself".

Given the gross inequalities of the day, More's idealized society required a formidable leap of the imagination. Now a small group of scientists from the unfashionable field of gerontology is urging us to consider a similarly radical leap. Today's utopian vision, they say, should be a society in which the elderly remain fit and healthy.

"People tend to associate old age with decrepitude and senility," says Rudi Westendorp, an expert on healthy ageing at Leiden University in the Netherlands. "But we have no reason to assume that weakness is inevitable in the old."

This message is a timely one, given the steadily ageing demographic in most industrialized nations. Advances in medicine have reduced the terrible toll of infant mortality caused by infectious disease, and more

recently have started to overcome the killers of middle age, such as heart disease and cancer. Life expectancy is rising, and shows no sign of levelling off (see 'Three score years and ten ... and more', opposite). At the same time, birth rates are falling, raising the spectre of societies that are unable to generate the cash to pay for the care of their elderly. There's an economic incentive, as well as a humanitarian one, for trying to break the link between old age and ill health.

Healthy outlook

At present, research into geriatric medicine is dominated by attempts to treat conditions, such as Alzheimer's disease, that mostly afflict the elderly. But with the demographic time bomb ticking loudly, politicians and research leaders are beginning to recognize the need to investigate the genetic and environmental factors that allow some people to remain healthy and active into their eighties, nineties and beyond.

The European Commission is now launching what will be the largest-ever study of the extremely old. The Genetics of Healthy Ageing (GEHA) project will gather genetic,

health and lifestyle information on 2,800 pairs of siblings who are more than 90 years old, who will be compared with the same number of younger controls. The idea is to find out what genes the elderly brothers and sisters have in common, and which of these occur more frequently than in the general population.

And later this month, Germany's Max Planck Society is expected to approve the establishment of a new institute to study the factors that influence longevity in laboratory animals. Both approaches will be needed, experts say, if researchers are to devise drugs and other interventions that will promote healthy old age.

Research on model organisms such as the nematode worm *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster* has already shown that genes can influence lifespan. In one of the earliest, and most dramatic, demonstrations, researchers led by Cynthia Kenyon of the University of California, San Francisco, showed that mutations reducing the activity of a gene called *daf-2*, which among other effects seem to slow metabolism, could double the lifespan of

*C. elegans*¹. Since then, studies in animal models have identified dozens of other genes that influence longevity².

But genes are only part of the story — as research on human populations has shown. Studies on twins indicate that only 25% of the variation in life span³, and half of the variation in cognitive function late in life⁴, can be attributed to genetic differences.

The rest is down to environmental and behavioural influences. So what can we do to increase our chance of living a long and healthy life? Not smoking is the most obvious piece of advice; moderating our consumption of saturated fats and alcohol is also likely to pay dividends. In rodents, diets that are extremely low in calories extend lifespan⁵. But it's not clear whether the same applies to people⁶ — and who, in any case, wants to live on a near-starvation diet?

Indulgence in tobacco, drink and food can't fully explain the huge differences in length and quality of later life seen in most human populations. So if we are ever to devise individually tailored interventions to extend healthy lifespan, we need a more systematic view of the interactions between genetic and environmental factors that influence longevity. "There are lots of opinions and theories about how lifespan is determined at the molecular level, but few facts," says developmental biologist Peter Gruss, president of the Max Planck Society.

Systems failure

Most gerontologists believe that the deterioration associated with old age is caused by failures in multiple physiological systems, resulting from a variety of physical stresses. "We now think of ageing as strongly influenced by the way we lead our lives — an accumulation of different faults, unrepaired wear-and-tear, whose trajectory can perhaps be predicted from birth," says Tom Kirkwood, a gerontologist at the University of Newcastle upon Tyne, UK.

Ideally, human studies would follow the same design as those of model organisms, studying large numbers of individuals throughout their lives. But this is impractical, so researchers have typically either recruited cohorts of people who have lived to an unusually ripe age, or turned to historical data.

Archived information is often incomplete or unreliable, but Westendorp and Kirkwood have made use of the highly accurate and extensive records of births and deaths kept for the British aristocracy, showing that the longest-lived women tended to have fewer children⁷.

This fits well with results from fruitflies, in which selective breeding for extended lifespan results in reduced fertility⁸ — suggesting that there is a fundamental trade-off between reproduction and longevity. This is exactly what evolutionary biologists would

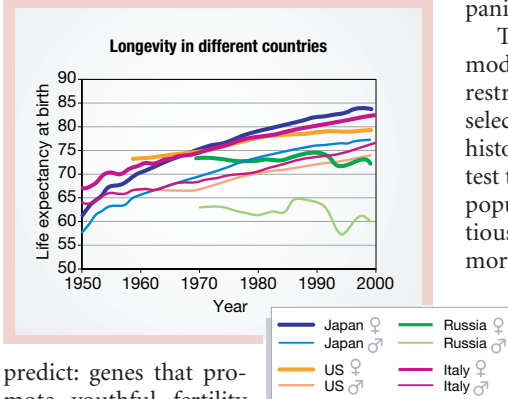
Three score years and ten ... and more

In the developed world, life span is continuing to rise (see Chart) — and given that the human record for longevity is 122 years, there's no suggestion that this trend should plateau any time soon.

Exceptional longevity tends to run in families, although whether this is primarily a result of shared genes or shared environment is unclear. And if you want to live a long life, it helps to be born female — or at least not to indulge in typically reckless male behaviour. "We reckon that one extra year is biologically inborn, and five or six years can be put down to differences in social behaviour between the sexes," says Rembrandt Scholz, a biomathematician at the Max Planck Institute for Demographic Research in Rostock, Germany. "Men take more risks — they drive cars faster, smoke more."

It also helps to be relatively wealthy. Trends in Russia, where male life expectancy tumbled during the hardships that followed the fall of communism, provide a stark reminder of the link between longevity and economic prosperity.

Alison Abbott and Anna Wellmann



predict: genes that promote youthful fertility should be favoured by natural selection, even if they cause problems in later life, after an animal's reproductive peak⁹.

The trade-off is probably the result of a division of metabolic investment between reproduction and the repair mechanisms needed to fix general wear and tear. Genes in the *daf* family, for instance, help to regulate nematodes' cellular repair processes and



Mutations in the gene *daf-2* can double the lifespan of the nematode *Caenorhabditis elegans*.

signalling pathways that help to protect against stress-related damage².

A similar trade-off may operate in the immune system — with genes that promote our ability to ward off infections as youngsters being detrimental to health in old age. Claudio Franceschi of the University of Bologna in Italy has studied genes regulating the production of molecules called cytokines, biochemical messengers that help to regulate our immune responses. He has found that centenarians tend to have genes that favour the production of cytokines that damp down inflammation over those that promote it¹⁰.

Inflammatory remarks

Franceschi has developed a theory that he calls 'inflamm-ageing', pointing out that many of the conditions that afflict the elderly are associated with inflammatory responses. "Low-level chronic inflammation is associated with all the classic diseases of old age, including Alzheimer's," he argues. High levels of inflammatory cytokines are also associated with the muscle weakness — known as sarcopenia — that often accompanies old age¹¹.

The inflammatory responses that in modern industrialized societies seem to restrict healthy longevity may have been selected for over most of our evolutionary history, when infections were rampant. To test this idea, Westendorp is now studying a population in northeast Ghana, where infectious diseases still cause appalling infant mortality. He is collecting mouth swab samples from newborn babies to determine whether the possession of genes for inflammatory cytokines correlates with survival. Westendorp is also studying cytokine genes in the oldest members of the same population, who should, if Franceschi's theory is correct, show less strongly inflammatory cytokine profiles.

Researchers are already planning clinical trials to see whether anti-inflammatory drugs can help to prevent diseases of old age. Some involve aspirin; others will test newer drugs that are more specifically targeted at the brain or other organs.

Franceschi's hunt for genes common to healthy Italian centenarians extends beyond the immune system. So far, he has looked at about 70 candidates, some of which are related to genes that have already been linked to longevity in model organisms; others were selected based on hypotheses about the likely causes of human ageing. Already, Franceschi has found several intriguing associations between extreme longevity and particular mutations, including some in genes that influence energy metabolism¹².

Other studies have looked at associations between longevity and genes involved in the metabolism of cholesterol, or its transport in

the blood. The latter involves a series of lipid-protein complexes, or lipoproteins, of various sizes and densities. Blood cholesterol is typically measured in association with high-density lipoproteins (HDL) — 'good' cholesterol — and with low-density lipoproteins (LDL) — the 'bad' cholesterol that gets deposited in blood vessels, leading to cardiovascular disease.

Franceschi's team, for instance, has shown that certain variants of the gene for an enzyme that protects against the damaging effects of LDL cholesterol are more common in centenarians than in the general population¹³.

One of the strongest associations is for variants of the gene for apolipoprotein E, a key component of the various cholesterol-carrying lipoprotein complexes. Those with at least one copy of the *ApoE4* form of the gene have a higher risk of cardiovascular disease and Alzheimer's disease in middle age. Centenarians are half as likely to carry this form as the general population, and are more likely to carry the *ApoE2* variant¹⁴, which may protect against these diseases.

True or false?

One danger is that candidate gene studies may be throwing up false positive results, if the cohorts of long-lived people are not well matched genetically to control groups from the general population. To get round this problem, Nir Barzilai of the Albert Einstein College of Medicine in New York has studied Ashkenazi Jews, who are relatively genetically homogeneous. This makes it more likely that genetic differences between controls and the very old really are associated with longevity.

Barzilai measured the size of HDL and LDL complexes in the exceptionally old and their children, and compared them with control subjects who were the same age as the children. Larger lipoprotein complexes — which are believed to reduce the risk of cardiovascular disease — were more common in the exceptionally old and their offspring¹⁵.

Another approach to avoid false positives is to scan the entire genome for genes that promote healthy ageing without any prior hypothesis of what they might be. This can be done by analysing single-letter variations in the genetic code known as SNPs, or single nucleotide polymorphisms. If exceptionally old people share particular SNP variants, this suggests that a gene influencing longevity lies nearby.

Researchers led by Tom Perls, a specialist in geriatric medicine at Boston University School of Medicine, have used this approach to compare 137 sets of siblings, all more than 90 years old, with a control group from the general population. SNP analysis pinned down the location of one longevity gene to

"We'd like to understand how some people manage to avoid the diseases of ageing, so that we can help others who may not have a natural advantage to do the same."



Record holder: Jeanne Calment, who died in 1997, lived to be 122 years and 164 days old.

a small region of chromosome 4 (ref. 16). A subsequent detailed scan, led by researchers from Elixir Pharmaceuticals, a company in Cambridge, Massachusetts, that is especially interested in diseases of old age, implicated variants of a gene called *MTP*, thought to be involved in the production of lipoproteins¹⁷.

The protein encoded by *MTP* has already been eyed up as a potential new target for cholesterol-lowering drugs¹⁸. Although the first generation of compounds was too toxic, gerontologists are excited about the possibility of extending healthy life by developing drugs to hit this target — or other parts of the pathways for metabolizing or transporting cholesterol.

The collaboration between Perls and Elixir will be dwarfed by the European GEHA study. In addition to their SNP analysis across 2,800 pairs of elderly siblings, GEHA researchers hope to collaborate with scientists at the Institute of Population Research at Peking University in Beijing. This would allow them to incorporate blood samples from about 3,000 ageing Chinese. "The more samples the better," says Bard Geesaman, Elixir's vice-president of medical development. "I'm not aware of any other study that is so ambitious."

The results won't be available for at least five years. But Westendorp predicts that GEHA will allow symptoms long regarded as

an unavoidable consequence of ageing, such as sarcopenia, to be elevated to the status of treatable diseases, with known genetic and environmental causes. "We'd like to understand in detail how some people manage to avoid the diseases of ageing, so that we can help others who may not have a natural genetic advantage to do the same," agrees Geesaman.

It sounds like a utopian future, indeed. Let's hope that it proves to be more readily attainable than Thomas More's egalitarian social vision.

Alison Abbott is Nature's senior European correspondent.

1. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. *Nature* **366**, 461–464 (1993).
2. Guarente, L. & Kenyon, C. *Nature* **408**, 255–262 (2000).
3. Skytthe, A. *et al. Twin Res.* **6**, 448–454 (2003).
4. McGue, M. & Christensen, K. *Exp. Aging Res.* **28**, 435–451 (2002).
5. Weindruch, R. & Walford, R. L. *The Retardation of Aging and Disease by Dietary Restriction* (Charles C. Thomas, Springfield, IL, 1988).
6. Heilbronn, L. K. & Ravussin, E. *Am. J. Clin. Nutr.* **78**, 361–369 (2003).
7. Westendorp, R. G. J. & Kirkwood, T. B. L. *Nature* **396**, 743–746 (1998).
8. Zwaan, B. J., Bijlsma, R. & Hoekstra, R. F. *Evolution* **49**, 649–659 (1995).
9. Kirkwood, T. B. L. & Austad, S. N. *Nature* **408**, 233–238 (2000).
10. Franceschi, C. & Bonafè, M. *Biochem. Soc. Trans.* **31**, 457–461 (2003).
11. Visser, M. *et al. J. Gerontol.* **57**, M326–M332 (2002).
12. Rose, G. *et al. Exp. Gerontol.* **38**, 1065–1070 (2003).
13. Bonafè, M. *et al. Eur. J. Hum. Genet.* **10**, 292–296 (2002).
14. Schachter, F. *et al. Nature Genet.* **6**, 29–32 (1994).
15. Barzilai, N. *et al. J. Am. Med. Assoc.* **290**, 2030–2040 (2003).
16. Puca, A. A. *et al. Proc. Natl Acad. Sci. USA* **98**, 10505–10508 (2001).
17. Geesaman, B. J. *et al. Proc. Natl Acad. Sci. USA* **100**, 14115–14120 (2003).
18. Jamil, H. *et al. J. Lipid Res.* **39**, 1448–1454 (1998).

Time to choose the right site for a fusion reactor

Europe is ready, willing and able to host the ITER, so let's get on and meet the challenge.

Sir — Your Editorial “Time for Japan to shine?” (*Nature* 427, 763; 2004) clearly presents the necessity of building ITER, formerly the International Thermonuclear Experimental Reactor. But I must say — as a European negotiator in the ITER talks — that, in trying not to be Euro-parochial, you do injustice to the role played by Europe in controlled fusion and in ITER.

Indeed, there is no mention of Europe's uncontested leading role in fusion, nor of the fact that data from the largest and best-performing machine, the Joint European Torus (JET) — which is above par — proved essential in designing ITER and in creating our confidence that ITER will meet its aims. The European Union (EU), more than any other partner, has provided constant, substantial support to the ITER project since its inception. The remarkable support given by Japan should also be mentioned at this point, although in terms of financial and human resources it has taken place at a significantly lower level.

Since July 2001, the EU has kept the project alive by spending €160 million

(US\$197 million), at a time when the United States had left the project. Your assertion that political and public support for ITER is less than whole-hearted in Europe does not rest on facts. The European Council of Ministers decided unanimously to present Cadarache, France, as the site, with reference to the financial estimate of costs made by the European Commission. The people living around Cadarache support ITER, and the region of Provence-Alpes-Côte d'Azur unanimously agreed to pay €447 million towards its costs. Although you say that Japan has a greater commitment to future energy sources, the legal framework for licensing ITER does not yet exist in Japan — whereas the licensing procedure has already started in Cadarache.

Nature correctly states that, if Europe's case is technically strongest, then Japan's compensation should include international contribution to an upgrade of the JT-60 tokamak to achieve critical science on the way to Demo, the engineering prototype reactor. But the EU negotiators deplore the

fact that the United States and Japan refuse a direct and objective comparison of the technical assets of the European and Japanese sites. Is this because Cadarache meets most of the nine criteria better?

For the above reasons, we feel that the ITER partners should now quickly decide to locate ITER in Cadarache, and that this decision should be accompanied by a commitment at the highest political level to implement a broader approach to fusion energy, well balanced between the needs of the programme and the capabilities of the partners. The world programme could include, in addition to ITER, the technology programme required for a commercial reactor (for example, a material-testing facility) and joint exploitation of ‘satellite’ tokamak facilities. This coherent approach would be the best way of getting on as quickly as possible with one of the most important world challenges of this century.

Paul Vandenplas

Consultative Committee Euratom-Fusion, and Ecole Royale Militaire, 30 avenue de la Renaissance, 1000 Brussels, Belgium

US science has never been more coherent

Sir — Your Editorial “Budget let-downs” (*Nature* 427, 571; 2004) paints a picture of US science funding that differs astoundingly from reality. The trajectory of R&D growth that you call “lacklustre” has risen more rapidly during the past four years than at any time in the past three decades: an average of 10% per year under the current administration. At US\$132 billion, the proposed 2005 R&D budget is at an all-time high and consumes a greater fraction of the domestic discretionary budget than at any time since the height of the Apollo space programme.

The proposals for priority programmes at the National Institutes of Health, the National Science Foundation, NASA and other agencies are above current and anticipated inflation levels and are well above the 0.5% increase for other parts of the non-defence discretionary budget. Priorities are well defined, established with wide interagency planning, and generally supported across agencies.

The proposed reorientation of NASA integrates robotic and human exploration, strengthens the scientific rationale for human spaceflight, and sets long-term budget guidelines that will protect NASA

science from overruns elsewhere. The Hubble Telescope decision is not connected with this reorientation, but strongly related to safety issues identified by the Columbia Accident Investigation Board.

The United States leads the world in providing resources to pursue critical research in areas such as environment, energy and global health. Some programmes have been reduced or eliminated on the basis of their performance record; others have been enhanced. Investments in areas of research related to domestic security, including emerging infectious diseases, belie your puzzling statement that “pressing scientific challenges ... evidently cannot attract sufficient resources.”

Equally puzzling is your failure to discover direction and coordination in a budget shaped by specific interagency initiatives for security, space exploration, climate change, nanotechnology, information technology and energy initiatives. These are identified so explicitly in the budget materials that one marvels at your statement: “No sign of that this year”. On the contrary, the coherence and strength of US science has never been greater or more productive.

John H. Marburger

Office of Science and Technology Policy, Executive Office of the President, Washington, DC 20502, USA

More to consider about European research body

Sir — The Royal Society's working paper on the European science base, discussed in your News story “Europe warned against research council” (*Nature* 427, 184; 2004), is much more positive about the creation of a proposed European Research Council (ERC) than you suggest. We do express reservations about certain proposals, but this does not mean that a more focused and prioritized version would not benefit research within the European Union (EU).

One of the main reasons that an ERC has been proposed is to increase spending on research and development in the EU, to close the gap with the United States. However, 90% of this gap is due to differences in business expenditure, so the potential creation of an ERC must not divert attention from this problem.

Our paper (www.royalsoc.ac.uk/policy) highlights other issues, including gaps in knowledge, that should be considered. We are studying proposals in the final Mayor report (www.ercexpertgroup.org), released on 15 December 2003, and hope that they will stimulate an informed wider debate.

Julia Higgins

The Royal Society, 6–9 Carlton House Terrace, London SW1Y 5AG, UK

High-tech cluster bombs

Why successful biotech hubs are the exception, not the rule.

Scott Wallsten

The first economic-development fad of the twenty-first century has arrived — Silicon Valley is out, biotechnology is in. The dotcom bust made Internet-led growth less attractive, but dreams of building a vibrant high-tech ‘cluster’ live on. Around the world, politicians wishing to replicate the biotech success stories seen in areas such as San Diego, California, are pouring taxpayer’s money into new initiatives. They imagine that luring the right businesses and research organizations to their backyard will set in motion a virtuous circle, attracting highly educated people to the area who will start more companies, which, in turn, will attract more people who set up even more companies.

There is just one small problem. Despite the promise of thousands of new jobs, cluster-development schemes rarely work. Success stories are the exception, not the rule. San Diego is a real biotech hub¹, as are Boston’s Route 128 and North Carolina’s Research Triangle Park, where thousands of scientists and engineers busily work on leafy corporate campuses. Yet nobody really knows how to start a successful cluster from scratch. As a result, money spent trying to build one is usually wasted.

Although the desire to set up a biotech cluster is relatively new, industry clusters have been around for some time². In the early nineteenth century, US manufacturing was concentrated in small parts of the northeast and midwest — shoes were produced in Massachusetts, and rubber in Akron, Ohio. Carpet producers still congregate in Dalton, Georgia, and furniture makers can be found in High Point, North Carolina.

Why do clusters thrive? In 1920, the economist Alfred Marshall suggested three common-sense reasons³: benefits of a pooled labour supply; access to specialized inputs; and information flow between people and institutions. These largely hold true today, although the type of labour and expertise vary by industry.

For example, high-tech regions have lots of venture capital, active universities and productive interactions between academic and industry scientists. Research by economists⁴ suggests that knowledge and information will ‘spill over’ from universities to benefit the surrounding communities. But politicians and their consultants have come to see such features as ingredients that, when combined, constitute a sure-fire recipe for high-tech greatness.



Reap what you sow? Not all biotech parks will be as prosperous as San Diego, despite investments.

It is tempting to think that all you need is a catalyst — tax breaks, for instance, or a high-profile institute that moves to the area — to get the process started. After all, Research Triangle Park, the grandfather of research parks established in 1959, became a success after IBM and the National Institute of Environmental Health Sciences set up shop there in the 1960s. Today, more than 38,000 people work in the 131 organizations located in the park.

With success stories such as this in mind, politicians are eager to push their own development ideas. The governor of Florida, Jeb Bush, recently said that giving the Scripps Research Institute in La Jolla, California \$310 million in state money and some \$200 million in local taxes to build a biomedical research facility in Palm Beach County was “the opportunity of a lifetime”.

Increasingly in the United States, cities, counties and states answer the call and put up taxpayer dollars for these high-tech dreams. Sometimes they think that direct subsidies, such as the Scripps deal, will do the trick⁵. At other times they designate an area as a research park, build some facilities and sweeten the pot with tax incentives to encourage businesses to move in.

In Baltimore, Maryland, officials are hoping that a biotech centre in a run-down section of town will lead to a local renaissance. Likewise, the secretary of commerce for Virginia noted that a new facility being built in Loudon County by the Howard

Hughes Medical Institute would be “an extraordinary windfall for Virginia”. State officials encouraged the county to give the institute a break on local property taxes to make the state seem more hospitable to other biotech companies.

Over the years, such activities have seen research parks sprout up like mushrooms around the world. The International Association of Science Parks lists members in more than 80 countries, and the United Nations estimates there are more than 400 parks worldwide. The Association of University Research Parks counts about 180 parks in the United States and has more than 200 members (not all members are parks, many are developers).

Are these initiatives based more on fantasy than fact? Consider a consultant’s report on the predicted economic impact of the Scripps deal in Florida. The report assumes that Scripps’ east-coast growth will mirror its San Diego experience, employing 2,800 people within 15 years and generating more than 44,000 jobs in the state.

Unlike Florida, officials in Baltimore at least seem to acknowledge the risks of their plan to attract biotech companies. And unlike most deals, Baltimore’s plan also includes substantial relocation payments to people being displaced from a poor neighbourhood. So at a minimum (and ignoring the fact that some residents might not wish to move), Baltimore may have found a way to help low-income families seek greener

R. TOMLINSON/ROBERT HARDING PICTURE LIBRARY/ALAMY

pastures. But predictions of 8,000 biotech-related jobs within ten years will probably turn out to be science fiction.

Consultants' reports rarely mention failed projects — Florida taxpayers might have been less inclined to hand over half-a-billion dollars if the job-growth predictions had been based on, say, the experience of Prince George's County in Maryland or San Antonio instead of San Diego.

The Maryland Science and Technology Center opened with great fanfare in the late 1980s, promising to employ 12,000 people and generate an additional 25,000 jobs in Prince George's county. Those dreams never came to fruition, as by the summer of 2003 the park's developer had built only 240,000 square feet of office space⁶. A local council member recently called the centre a failure, and the city plans to rezone the site to allow residential construction. The state wants a refund on some of the millions of dollars it invested in the site's infrastructure.

San Antonio broke ground on its Texas Research Park in 1987, with claims that it would lead to 30,000 jobs on the park and another 100,000 spin-off jobs within 30 years. Thirty years have not yet elapsed, but with just 15 companies and about 300 jobs, it does not look too promising.

Both the successes and failures make for fascinating anecdotes, but they cannot answer a fundamental question: do research parks have an effect on regional development? Recently, I set out to investigate this issue empirically⁷.

I assembled data on the number of jobs by industry (from the US Census) and amount of venture capital (from Venture Economics) for each county in the United States covering more than a decade. I then matched counties that established research parks in that time period (the 'experimental' counties) with similar counties that did not undertake such a venture ('control' counties). This method meant that the experimental group only included counties that established a park in that time period, thus excluding the early success stories, such as Research Triangle Park, as well as older failures. Matching counties was imprecise at best, but I considered counties to be similar if the number of high-tech jobs, venture capital and population were comparable the year before the experimental county set up its park.

I then compared job and venture-capital growth in the experimental and control counties before and after the parks were established. It turns out that the number of jobs or the amount of new venture capital grew no more in the counties that established research



K.C. Nicolaou (left) and Richard Lerner of the Scripps Research Institute show Florida Lieutenant Governor Toni Jennings the San Diego site.

parks than in counties that did not. In other words, the data provide no evidence that, on average, setting up a research park made any difference to the region's economic growth.

I also undertook an econometric analysis that controlled for income, large firms in the county (that could help to attract or spin off smaller firms), research spending by universities, government jobs, a time trend and other factors specific to a given county. This analysis reached the same conclusion. On average, there was no evidence that building a research park made a difference to regional development. If you judge these job-creation schemes not even by their own lofty goals, but by the simple expectation of driving regional growth, there is little evidence that they work.

Some may argue that not every cluster can be a San Diego — you cannot judge them all by the same criteria. This is exactly right. Almost by definition, a given industry can cluster in only a few places. But those who promote biotech initiatives never say that they want to be a second-rate hub.

How do we explain the exceptions? Some, such as Research Triangle Park, had the advantage of being first. Others, such as Silicon Valley, benefited from a unique set of entrepreneurial and venture capital networks that developed symbiotically with local research institutions like Stanford University. Government policies that encourage entrepreneurs and promote a sound investment climate through, for example, good schools and infrastructure, also help. However, governments — both local and national — are not usually good at 'picking winners' when it comes to predicting which technology will drive future growth. It is even less likely that a given region will succeed if everyone is chasing the same dream.

The definition of success itself should be closely examined. If a park claims 100 new

companies, is that a success? To the park's promoters it is. But from the region's perspective it is only a success if those companies did not simply 'cross the street' to take advantage of tax credits inside the park boundaries. Even if the businesses are new to the region, it is only a success if the benefits of luring them there outweigh the costs. Viewed from a broader perspective, Floridians will celebrate if companies move from Massachusetts to set up shop near Scripps. But globally, there is no net benefit. It is just a transfer — of companies from one place to another and of money from taxpayers to mainly wealthy organizations and developers.

The impact on the scientific community is harder to predict.

New initiatives presumably mean new job opportunities and potentially higher wages, which could, in turn, attract more people to the field. But if biotech growth is not sustainable then — as with the dotcom experience — that influx of people could turn into an oversupply. More generally, failed clusters across the country are unlikely to do much to enhance biotech's public image.

The obsession with becoming the next biotech hub will fade in time, just as dotcom dreams did. But research parks and similar schemes to promote high-tech development are likely to continue to be irresistible to nearly everyone involved. Businesses — from high-tech companies to developers — that get money from these schemes are happy with their windfalls. Politicians are happy to hand out pork while looking like visionaries.

These are tough forces to overcome, but there are things we can do. First, economists and other social scientists should put more effort into understanding regional economies and finding ways to help regions improve on their strengths. Second, citizens should demand honest evaluations of cluster initiatives, instead of glowing reports from the politicians' hired guns. Some high-tech development schemes may prove to be good investments, but the vast majority should remain in the realm of science fiction, where they belong. ■

Scott Wallsten is a fellow at the AEI-Brookings Joint Center for Regulatory Studies, 1150 17th St. NW, Washington DC 20036, USA, and a resident scholar at the American Enterprise Institute.

1. *Nature* **426**, 689–721 (2003).
2. Krugman, P. *Geography and Trade* (MIT Press, Cambridge, Massachusetts, 1991).
3. Marshall, A. *Industry and Trade* (Macmillan, London, 1920).
4. Jaffe, A. *The Am. Econ. Rev.* **79**, 957–970 (1989).
5. *Nature* **426**, 4 (2003).
6. Wilen, J. *Baltimore Business Journal* (13 June 2003).
7. Wallsten, S. *Do Science Parks Generate Regional Economic Growth?* (AEI-Brookings Working Paper, Washington DC, 2004).

Why we did it

An account of the driving forces behind the unfolding of human civilization.

A Brief History of the Human Race

by Michael Cook

W. W. Norton: 2003. 384 pp. \$26.95

Granta: 2004. £20

Melvin Konner

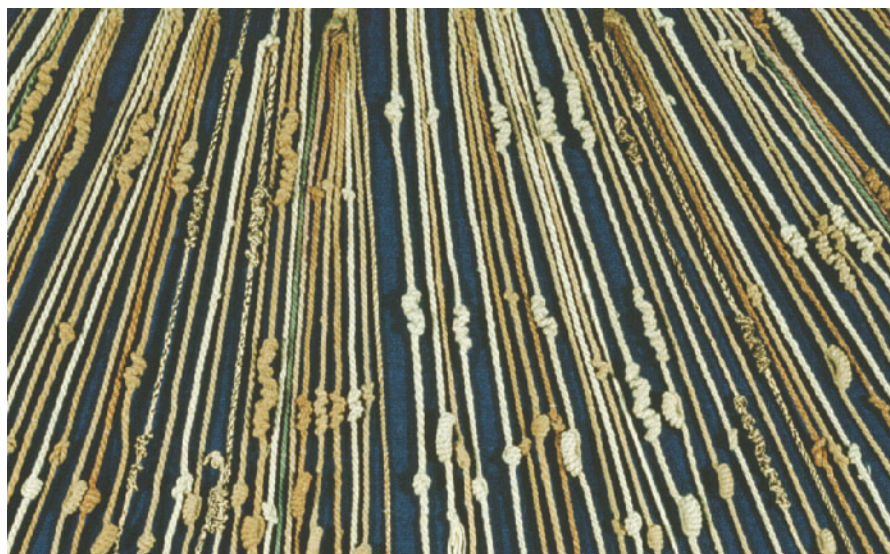
Ask the average historian whether a science of history is possible and you are likely to get a condescending smile. History, they assure us, belongs to the humanities; it is not even a social, much less a natural, science. Historiography is an intensely empirical exercise, but it deals with the facts of one time and place, or at most the record of how events have unfolded on a larger scale. It is not for nothing that the field's name contains the word 'story'; the story must be as true as it can be, but do not expect it to yield any patterns.

Attempts at pattern-finding during the nineteenth and early twentieth centuries — by Karl Marx, Oswald Spengler and Arnold Toynbee, for example — are seen as cautionary failures. There are world-systems theorists today, but they are few and marginalized. There are cliometricians, who present the particulars in quantitative form, but they are unlikely to generalize. Most historians don't want to measure Clio, the muse of history — they want her to inspire them to tell the best and truest story.

Anthropologists are different. Some at least have always sought patterns, and in surveying the great array of cultural variation that archaeology has unearthed and ethnography revealed, they have found them. Patterns have cropped up in some basic textbooks for decades. For example, in *Culture, People, Nature* (Longman, 1993), Marvin Harris calls the New World "the Second Earth", and presents facts about the independent rise of agriculture, cities and empires as the results of a natural experiment: what happens if we start with hunter-gatherers and run the tape again? The results are remarkable. Do the Aztecs exactly replicate Sumer or Shang China? Of course not. But no one can come away from the comparison without the sense that, at the broadest level, there are laws of history.

In the modern period, not all historians have shied away from cause-and-effect relations. William McNeill, for example, in his respected *Plagues and Peoples* (Penguin, 1979), showed history to be shaped by epidemics, and his history of the West takes seriously the effects of demographic, environmental and technological forces.

Michael Cook takes this process further. *A Brief History of the Human Race* begins geologically, using data from the Greenland ice core to show that the past 12,000 years



Strands of time? The Incas used a knotted rope called a *quipu* to keep historical records.

have seen a climate both warmer and more stable than in the preceding 100,000 years at least. This, he argues, made history — the change from a hunter-gatherer world to ours — possible. He proceeds biologically, showing that those 100,000 years correspond roughly to the spread of modern humans out of Africa. His summary of the DNA findings on this process leads to a conclusion widely accepted in anthropology: the human species as we know it is quite new and therefore remarkably unified.

One consequence is that humans respond similarly to the challenges and chances that geology and geography afford them. All human groups have wanted to do well, and many have had the ambition to build, grow and change. But the way they did it depended on where they were. Cook devotes considerable attention to Earth history, using descriptions of the collision of tectonic plates, the smoothing over of mountains, and the rise and fall of oceans, to set the scene for each continent's historical drama. Next, he reviews the peopling of the regions, using the best available genetic and archaeological evidence. He goes on to explain many aspects of regional history by reference to each region's geology, ecology and climate.

Technological innovation, predictably, has a central role thereafter. Stone was replaced by copper, copper by bronze, and bronze by iron in many parts of the world because humans want to do things better, and if the environment or trade networks afford them the opportunity, they will discover and apply new technological knowledge. Anomalies in this regard — some New Guinea groups who never left stone behind, or the absence of large domesticated animals

in North American civilization — are attributed to limited ecological opportunities or geographical barriers to communication with other groups.

The book is anthropological in another sense: Cook presents vivid accounts of an important aspect of culture for each region in a way that is unusual in history books. Thus we find considerable detail on the Incan *quipu*, a record-keeping device in knotted rope; on the intricate clan and marriage rules of the Aranda hunter-gatherers of Australia; and on the sexually charged religious writings of India. These illustrate such general human tendencies as the need to keep account of things, the desire to order the social world, and the impulse to both express and control sexual drive.

By contrast, remarkably little attention is given to the past few centuries, which are covered in fewer than 30 pages. Perhaps this is appropriate. In a short history of the human species, many hard choices must be made, and Cook has focused on changes between the hunter-gatherer phase and the Industrial Revolution. Perhaps by that time the patterns were so well established that they could not be expected to change radically. But is hard to justify skipping the invention of democracy, our new power to destroy life on Earth, and the recent revolution in communication, while writing in detail about the *quipu* — however much it may warm an anthropologist's heart.

Another omission from the book — as from most history books — is any account of the human players in this great historical drama as individual members of a species, with shared tendencies and behaviours. The book is full of large-scale violence in the

service of competition and ambition, but it makes no explicit reference to the possibility that part of the explanation lies in human nature. There are many entertaining references to sex, but none to the reproductive imperative that makes sex so exceptionally interesting.

Still, this is a particularly good history for a scientist to read, devoting as it does much more time and effort to underlying material causes than has been traditional. It is an elegant, quick and engaging way to review what has happened in history, to learn much that is new, and to appreciate the past of the whole world, not just the West. It meets scientists almost halfway, trying to ground the events of history literally in the material facts of the planet. As Cook understands, the best and truest story of our experience on Earth is in part a scientific one. Clio, I think, is smiling. ■

Melvin Konner is in the Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. He is the author of *The Tangled Wing: Biological Constraints on the Human Spirit*, revised edition.

Sowing seeds of discontent

So Shall We Reap

by Colin Tudge

Allen Lane: 2003. 380 pp. £20

Anthony Trewavas

A common method of political and religious persuasion is to dwell on the virtues of belief for the follower and damnation for the unbeliever. Science writer Colin Tudge uses the same approach in this book, which is devoted to what he calls 'enlightened agriculture'. The prospects for humanity 'are somewhere between glorious and dire', we are told; glorious if you follow Tudge's proscriptions, hell if you don't. The book's subtitle hammers this home: "How everyone who is liable to be born in the next ten thousand years could eat very well indeed; and why, in practise, our immediate descendants are likely to be in serious trouble". And as if to emphasize the religious origin of such dichotomous futures for humans, the book is liberally spiced with biblical quotations.

Tudge regards food production as something that should be above the ordinary, grubby business of economics. He regards capitalism as acceptable provided it doesn't involve competition! He sees a future in which most of us return to the rustic idyll, happily tilling the land — although he does not say who will generate the cash to pay for education or health, or the other trinkets we have got used to and that people enjoy. Whether any of us want to return to that way



Cultivating a dream: do today's economic realities stop us returning to the rural idyll?

of life is not considered. And because it is generally agreed that there isn't much money to be made in farming, the prospects for most of us do not look good.

Antipathy to economics is common among those of Tudge's persuasion, but unless fine-sounding sentiments are properly underpinned with an understanding of economics, immense damage can result. It has been estimated that for every 1% increase in income, mortality is reduced by 0.05%. The converse is equally the case. Scepticism about the motives of large global agribusiness is reasonable, but assuming that they are populated with shadowy figures out to control the world's food supply is not. Most UK citizens (probably including Tudge) invest heavily in the success of such enterprises through pensions and other financial plans. From Malthus onwards, the history of agricultural prediction has been a history of failure. Tudge's poorly based views will probably fare no better.

In Western countries a few decades ago, agricultural policy was simple. Production was all that was needed, and objective knowledge (science) was wheeled in to ensure its success. But abundance has produced new problems. Food security is no longer an issue, although rapid global cooling could quickly push it up the agenda. Instead, agriculture and, in Britain at least, the inevitable intermingling with the environment, have become contentious moral affairs. These are now areas of subjective knowledge in which disagreement, which merely reflects individual taste, is inevitable. Tudge claims that biology is the basis of his book, but chapters covering such issues as morality, aesthetics, genetically modified (GM) organisms, cash and values belie the claim.

It is a pity that authors such as Tudge and most environmentalists do not talk to farmers, as a more realistic appraisal might then surface. Farmers could tell them about

responsible farming based on integrated management, conservation agriculture and animal-welfare principles, but also about the necessary business of running the farm at a profit. For the public, competition produces cheap fruit and vegetables, and by thus encouraging consumption has produced a healthier population with lower cancer rates. Tudge is more objective on organic farming and sees the regulations of this movement as dogma rather than common sense.

Tudge reserves his venom for GM crops, condemning the scientists who produce such "monstrosities" as obviously corrupt, as well as mad, bad and dangerous. I found this chapter to be a muddle of politics and naturalism, failing to adequately distinguish objective scientific knowledge from subjective assessments of Western agribusiness and nature.

Vitamin A deficiency in developing countries results in the premature death of about a million children a year and leaves another five million permanently blind. The primary reason for this situation, according to the World Health Organization, is poverty and ignorance about vitamins and diet. But Tudge claims instead that recent Western agricultural influences are the cause, as if these deficiencies did not happen in much earlier times. GM rice enriched with vitamin A to help counter this deficiency is a humanitarian scientific endeavour that demonstrates how valuable GM technology can be in improving life expectancy in the face of ignorance.

Tudge also fails to mention that both India and China now have proven examples of the benefits of GM crops to poor farmers. Herbicide-tolerant GM crops (produced by agribusiness) lead naturally to no-till agriculture, which has enormous environmental advantages over any kind of ploughed agriculture, including organic farming. These benefits are likewise not mentioned. GM

food vaccines? Guess what. Not mentioned. The list of omissions is very long. In a pluralist society you do not ban useful products.

Mad? Bad? Dangerous? The only real danger is those who use subjective ideology to corrupt good objective sense. ■

Anthony Trewavas is at the Institute of Cell and Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JH, UK.

The age-old problem of mortality

The Biology of Death: Origins of Mortality

by André Klarsfeld & Frédéric Revah
(transl. Lydia Brady)
Cornell University Press: 2003. 204 pp.
\$29.95

Caleb E. Finch

The Biology of Death provides an engaging travelogue for ageing and death, from ancient mythology, the swamps of vitalism and the higher roads of the 'copernican revolution' of experimental gerontology, to today's Société de Thanatologie with its fixation on the ontology of death. The authors, André Klarsfeld and Frédéric Revah, are widely read and show fine taste in their choice of anecdote and historical detail. I was delighted to learn more about the French anatomist Xavier Bichat and the Chevalier de Lamarck. And I hadn't heard before the anecdote about Henry Ford searching junkyards to find parts of his Model T that never wore out.

The book's style and level of explanation are highly suitable for a general audience. The broad overview of gerontological fact, fiction and theory is basically sound and provides a useful introduction to the perplexities of biological ageing. However, despite the authors' scientific credentials as experimental neurobiologists, some misunderstandings have crept in that will irritate the informed. It is a pity that these were not picked up in review by experts in biogerontology before the publication of either the French edition (Janvier, 2000) or this English translation.

The authors repeat a once common misunderstanding about the Hayflick limit that is still transmitted by the popular press: "These cells stopped dividing and died after having performed a fairly constant number of around fifty divisions." Cellular gerontologists, including Leonard Hayflick, now agree that 'senescent' cultures of fibroblasts can persist in stationary, but metabolically active, states for many months or even years (as do most of our brain and heart cells). Moreover, clonal analysis of daughter cells shows huge differences in the number of subsequent divisions, which are anything

Dress for DNA

Molecular genetics has proved a rich source of imagery and ideas for contemporary artists, as the DNA sequence/spinal-column dress shown here attests. It forms part of the collection *Primitive Streak*, a collaboration between designer Helen Storey and her sister Kate, a research biologist. Their work, and that of other artists inspired by the genetic revolution, is the subject of *The Molecular Gaze: Art in the Genetic Age* (Cold Spring Harbor Laboratory Press, 2004) by artist Suzanne Anker and the late Dorothy Nelkin, a social scientist. Anker and Nelkin seek to place these images in context within the history of art and to provide insight into the social and cultural meaning of the genetic revolution through the gaze of artists. They explore themes including the meaning of mutation and the blurring of species boundaries.

Mary Purton



but "fairly constant". The Hayflick limit therefore represents an ensemble average of very heterogeneous cell populations.

This intrinsic heterogeneity confounds interpretations of measurements made on extracts from millions of cells. For example, if gene A goes up and B goes down, relative to a younger stationary culture, it can not be known if A and B have changed in the same cell. The book also implies that the Hayflick limit can be used up during the lifespan. However, the doubling potential of skin fibroblasts does not alter as an adult ages.

The enormous variety of life histories and modes of death in different species of animals and plants intrigues the authors. They draw liberally from examples I collected over 30 years and mentioned in my own book *Longevity, Senescence, and the Genome* (University of Chicago Press, 1990), which they cite. I had hoped to find some mention of key examples that have since come to light, such as Justin Congdon's description of the negligible senescence of turtles. Dating of individual specimens of fish and other animals with radioisotopes has advanced far beyond the often uncertain counting of old growth rings (G. M. Cailliet *et al.* *Exp. Gerontol.* **36**, 739–764; 2001). Deep-sea fish in several families in addition to the scorpenids are now better documented to live beyond 100 years of age, including a gourmands' favourite, the orange roughy. But the anecdote about a 152-year-old sturgeon (*Comm. Fish. Rev.* **16**, 28; 1954 — authors not named), cited without reference, should not have been repeated without a fisherman's wink or a scholar's note — this was a caveat that got away.

The authors' exposition of the evolutionary theory of genetic determinants of longevity is excellent. They provide a lucid review of the classic models of reproductive schedules, which address adaptive trade-offs

between number of offspring, initial age of reproduction, reproductive senescence and age-group mortality, and give a basis for the evolutionary selection of lifespan. The authors then give a synthesis of more recent experimental work on fruitflies with field studies on both possums and guppies, which together show that predatory pressure can accelerate reproduction at the apparent expense of longevity. Curiously, though, they do not discuss genetics of longevity in human twins. Several twin studies have shown that lifespan has a relatively low heritability.

For a book on the biology of death it says little about the causes of death at advanced ages, which may be very different to those at younger 'old' ages. For example, for rats whose lifespan was extended by calorie restriction, autopsies revealed so little histopathology in 25% of rats that causes of death could not be assigned (I. Shimokawa *et al.* *J. Gerontol.* **48**, B27–32; 1993). It may be that, as the number of very old humans increases, we see cases of death caused by systems failure from multiple small lesions, any one or several of which would not be a sufficient cause on its own. A systems thanatology could be modelled to represent synergies of different levels of dysfunction, with statistical estimates of fluctuations from the norm.

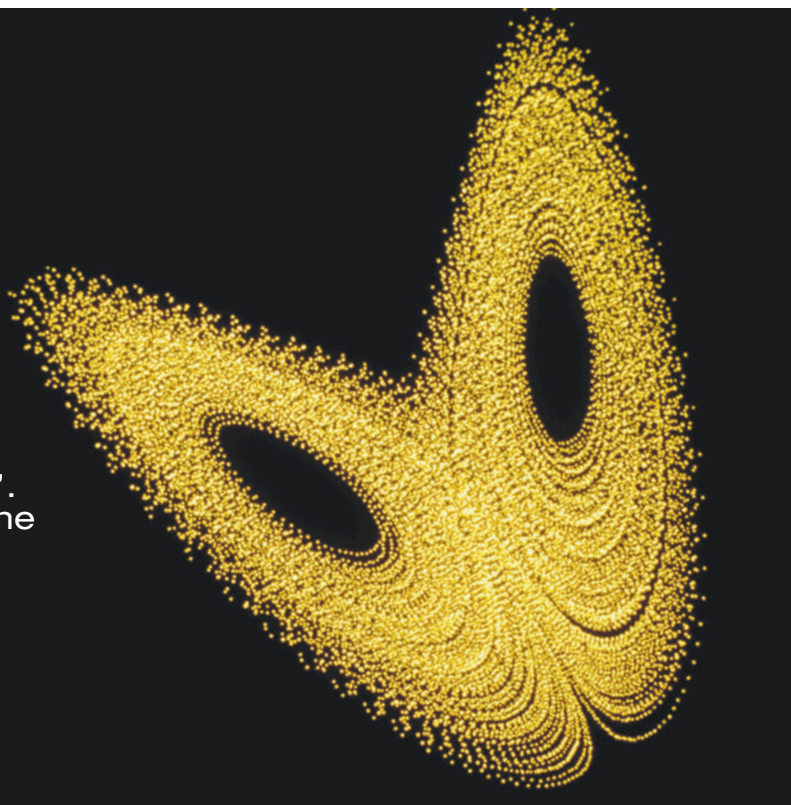
The Biology of Death belongs on the same shelf as other recent popular books on ageing, such as Hayflick's *How and Why We Age*, Steven Austad's *Why We Age* and Tom Kirkwood's *Time of Our Lives*. The remarkable progress being made throughout gerontology has provided scope for the frequent new accounts aimed at the general public. ■

Caleb Finch is at the Andrus Gerontology Center and Department of Biological Sciences, University of Southern California, Los Angeles, California 90089-0191, USA.

Hurricanes and butterflies

Thomas C. Halsey and Mogens H. Jensen

Chaotic systems can be characterized by the swirling patterns of 'strange attractors'. A powerful method to determine their behaviour has been validated for the most famous case, the Lorenz attractor.



PROF. E. LORENZ, PETER ARNOLD INC./SPL

Chaotic systems are famous for their sensitive dependence on initial conditions. Small changes in the original variables describing a system, induced perhaps by the flapping wings of a butterfly, can result in large changes in the outcomes — such as the magnitude of property damage in Florida, in the wake of a hurricane. History only happens once, so it would seem impossible to determine *a posteriori* that the path of a particular hurricane could have been altered if only the butterflies had been more cooperative.

However, writing in *Physical Review Letters*, Sam Gratrix and John N. Elgin¹ provide grounds for optimism: they have developed a powerful new method to determine from experimental observation of a system whether it is chaotic, and, if it is, what the precise quantitative nature of that chaos is. Their method is based on fractal geometry. Fractals are structures or curves that remain rough or heterogeneous on all length scales and are characterized by their 'fractal dimension'. The coastline of England is such a curve, as pointed out by Benoit Mandelbrot, because the coastline can be regarded as rough at least down to the scale of the individual sand grains on the beaches.

Dynamical systems are generally analysed in terms of their behaviour not in real space but in the phase space spanned by the variables describing the system — for a hurricane, these include values of atmospheric pressure, humidity and velocity, on a grid sufficiently fine to determine the future course of the hurricane. Depending on the number of such variables, the phase space for

Figure 1 The Lorenz attractor, tracing the phase-space trajectory for a simple model for atmospheric convection. The trajectory is quite densely bunched in some regions and is quite sparse in others, which makes this 'strange attractor' multifractal.

a particular system will have a corresponding number of dimensions. As they evolve, however, chaotic systems settle on some structure of a lower dimension than their corresponding phase space, structures called 'strange attractors'. These are fractals, and perhaps the most famous strange attractor is the Lorenz attractor (Fig. 1), discovered in 1963 by Edward Lorenz² for a model of atmospheric convection (hurricanes again).

But strange attractors are not simple fractals like the coast of England. They are 'multifractals', whose quantitative properties vary from point to point in an intricate manner, with the result that they are characterized by a range of fractal dimensions. Imagine that the strange attractor in Figure 1 is covered with small boxes; then consider what percentage of its time the system finds itself in any particular box. The distribution of times for the boxes is extremely wide — from boxes that are almost never visited by the system dynamics, to boxes in which it spends a disproportionately large amount of time. So it seems elementary to determine whether or not a particular dynamical system is chaotic: simply reconstruct its trajectory through phase space, cover that trajectory with boxes, measure the amount of time spent in each box, and then determine whether or not the multifractal structure you have computed is consistent with chaos.

This box-counting method to diagnose multifractality has been applied to a wide

variety of systems over the past 15 years. Many systems known on other grounds to be multifractal have had their multifractality confirmed in this way. Unfortunately, many systems known on other grounds not to be multifractal have also had their 'multifractality' confirmed in this way. Alas, box-counting proves little about multifractality, and much about the truth of a famous Mark Twain saying regarding statistics and lies.

Nevertheless, there is an alternative route to the determination of multifractal properties. Mathematicians know that the strange attractor can actually be constructed from the union of all periodic trajectories of a system, provided that trajectories of arbitrarily long periods are included (over a short observation time, these trajectories might not be obviously periodic, just as it takes at least 28 days' worth of observations to conclude that the Moon does, indeed, revolve around the Earth). Using an ingenious method to categorize these long trajectories, Gratrix and Elgin¹ have reconstructed in great detail both the Lorenz attractor and its multifractal properties.

Of course, in nature we rarely have the opportunity to observe a dynamical system for long enough to find enough trajectories to rebuild the strange attractor in such detail. However, the results from the periodic trajectory study can be compared with a much simpler approach, based on recurrence times. Recurrence times are simply defined:

consider a point on the attractor, and ask how long it will be before a trajectory starting at that point returns, not exactly to that point (as in the periodic trajectory calculation), but to within some certain distance of that point. Although a method was developed in the mid-1980s to find multifractal properties based on the properties of the recurrence times³, this method had not been applied to strange attractors, and had not been verified against more rigorous methods (an important step, given the unfortunate history of box-counting).

These gaps have now been filled by Gratrix and Elgin¹, by developing the recurrence-time method for the Lorenz attractor and by verifying it against the periodic trajectory method. Because calculations based on recurrence times should be relatively straightforward for experimentalists, and as we now have reason to believe that they will be more reliable than box-counting results, we can confidently await a new series of experimental demonstrations of the chaotic properties of a variety of natural systems.

But will this solve the problem of the butterfly and the hurricane? The Lorenz attractor lives in a three-dimensional phase space; a hurricane lives in a phase space with an enormous number of dimensions. After several decades of work on chaos, we still do not understand the extent to which systems with such large numbers of degrees of freedom (typically turbulent systems) can be understood using the same concepts as for chaotic systems, which are relatively simple in comparison. So if a hurricane destroys your beach house, the verdict against the butterfly is: not proven. ■

Thomas C. Halsey is with ExxonMobil Research and Engineering, Route 22E, Annandale, New Jersey 08801, USA.

e-mail: thomas.c.halsey@exxonmobil.com

Mogens H. Jensen is at the Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark.

e-mail: mhjensen@nbi.dk

1. Gratrix, S. & Elgin, J. N. *Phys. Rev. Lett.* **92**, 014101 (2004).
2. Lorenz, E. N. *J. Atmos. Sci.* **20**, 130–141 (1963).
3. Jensen, M. H., Kadanoff, L. P., Libchaber, A., Procaccia, I. & Stavans, J. *Phys. Rev. Lett.* **55**, 2798–2801 (1985).

Human longevity

The grandmother effect

Kristen Hawkes

Why do women live long past the age of child-bearing? Contrary to common wisdom, this phenomenon is not new, and is not due to support for the elderly. Rather, grannies have a lot to offer their grandchildren.

Those who think postmenopausal women make little difference in the story of human populations will be surprised by the report of Lahdenperä and colleagues on page 178 of this issue¹. The authors have unearthed firm evidence in support of the 'grandmother hypothesis', according to which a grandmother has a decidedly beneficial effect on the reproductive success of her children and the survival of her grandchildren.

The question of human longevity has deeper evolutionary importance than many think. It is often assumed that the steady increase in life expectancy over the past century and a half² has resulted in a larger proportion of older people than ever before. But, until the past few decades, increases in life expectancy reflected reductions in infant and juvenile mortality, and made little difference to the fraction of women past child-bearing age. As shown in Fig. 1, it is levels of fertility, not life expectancy (mortality), that shift the proportion of elders in a population. Even when life expectancy is well below 40 years, most girls who survive childhood live past their child-bearing years. In both historical and hunter-gatherer populations, a third or more of women are usually beyond the age of 45.

This large proportion of older people has fundamental implications for all human social organizations. Its unusual character is highlighted by comparisons with other primates. For example, among our closest living relatives, chimpanzees, female fertility declines at about the same age as in people, from a peak before age 30 to virtually zero at age 45 (ref. 3). But chimpanzee survival rates fall along with fertility, so that in the wild less than 3% of the adults are over 45 (ref. 4).

We might assume that the large fraction of elders in human populations reflects a characteristically human social safety net. But natural selection generally favours the flow of help from older to younger kin (Fig. 2), so we should be sceptical that a species-wide pattern of care for older people explains human longevity⁵. Developments in evolutionary life-history theory suggest that, instead of help for older members of the population, it is help from postmenopausal grandmothers that accounts for the age structures of human societies.

Mammalian life histories fall along a fast-slow continuum⁶. At one end, maturation is quick, fertility is high and adults die young. At the other, maturity is delayed, reproduction is slow, and adults usually live long enough to grow old⁷. The most

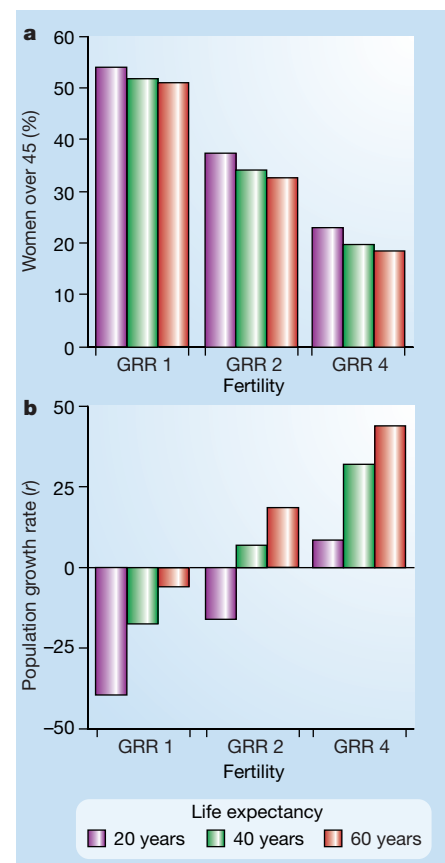


Figure 1 Population profiles for women — females over 15 years of age — for different variables. a, The proportion of women over 45 for different life expectancies (20, 40 and 60 years) and fertility levels (expressed as gross reproduction rate, GRR, which is equivalent to the average number of daughters for women who survive the fertile years). When life expectancies are the same, higher fertilities make younger cohorts larger and the proportion of elders smaller (compare all the purple bars, all the green bars, all the orange bars). When fertility levels are the same (each cluster of bars), the fraction of elders varies little even across a tripling of life expectancies. Life expectancies vary widely with differing levels of infant and juvenile mortality. The point to stress is that, at the same fertility level, populations with shorter life expectancies do not have fewer elders: in fact, they have slightly more women past fertile ages. b, Population growth rate (r = annual rate of growth/1,000) accompanying each combination of fertility and life expectancy. The very high proportions (> 50%) of women past 45 occur in sharply declining populations, and the very small proportions (< 20%) in swiftly growing populations. Growth rates that differ from zero cannot be sustained for long. (Data from ref. 17.)

successful model for explaining this cross-species variation⁸ shows that adult lifespans can determine the other life-history traits. The relationship between average adult lifespan and average age at maturity is much the same across the living primates, including humans⁹. Chimpanzees are at the slow

extreme; humans are even slower. From this perspective, the question to ask is, what caused the extension of lifespan in human evolution?

The answer may lie in grandmothers' contribution to childcare. Unlike other primates, including chimpanzees¹⁰, human children are unable to feed themselves when they reach weaning age. The foods we rely on are too difficult for young children to handle. This gives women whose fertility is ending (so they have no newborns of their own) an opportunity to influence the reproductive success of their daughters and survival of their grandchildren through assistance in food provisioning. In an ancestral population that was shifting from chimpanzee-like feeding to hard-to-handle foods¹¹, the more vigorous elder females could help more, thereby increasing the representation of their vigour in descendant generations, shifting rates of ageing¹², and lengthening average adult lifespans¹³.

This 'grandmother hypothesis' accounts for various similarities and differences between the life histories of human females and those of our nearest living relatives. We live longer, mature later and space births more closely during the child-bearing years, whereas our fertility declines and we reach menopause at similar ages¹⁴. But does the hypothesis stand up to scrutiny?

A few studies have hinted at the relationship between postmenopausal longevity and the welfare of a woman's descendants, but most have been concerned about the timing of menopause. Lahdenperä and colleagues¹ focus squarely on the longevity question. They have analysed multi-generation records from two eighteenth- and nineteenth-century populations, Finnish (slow-growing) and Canadian (fast-growing). They show that in both of these populations the duration of a woman's postmenopausal survival affects both the reproductive success of her children and the survival of her grandchildren.

In neither the Finnish nor Canadian populations are these effects due to secular changes in both family size and longevity, or geographical differences that might give the same relationship independent of grandmother effects. Other studies have sometimes identified associations between postmenopausal longevity and numbers of children¹⁵, and different grandmother effects through sons and daughters. Lahdenperä *et al.* find neither.

Finer-grained data for the Finnish population allow them to go further. Both sons and daughters who had a living mother past menopause had children sooner and at shorter intervals, and raised more of them to adulthood. These differences are independent of differences in wealth, and increased the longer the mothers lived. Fewer grandchildren were born if the grandmother was not living in the same village. Because post-



Figure 2 Helping hands for the young — Granny provides assistance.

menopausal mothers were at different ages when each of their children had children, Lahdenperä *et al.* could compare the success of children of the same mother depending on whether she was alive or dead, improving the likelihood that differences are due to her help. Grandmothers' effect on the survival of grandchildren did not begin until the children reached weaning age, another detail indicating that the investigators were directly measuring her help. By controlling so many of the likely confounding factors, Lahdenperä and colleagues show that the family assistance provided by grandmothers is a central determinant of our longevity.

We age slowly. Physiological mechanisms must underlie that¹⁶. Accumulating evidence (referenced in ref. 13) shows that humans

allocate more to cell and molecular maintenance and repair than do our nearest primate relatives. The grandmother hypothesis attributes our slow ageing to the help that older females can give their descendants¹²; females ageing more slowly in physiological systems other than their ovaries could help more. Questions about male life histories involve different trade-offs, but this hypothesis about shifting rates of ageing has implications for longevity in both sexes. Daughters and sons both inherit genes affecting levels of cellular maintenance and repair from their mothers. While others work to unravel these mechanisms, Lahdenperä and colleagues have found evidence that can help to explain why they are set differently in us than in our nearest non-human kin. ■

Kristen Hawkes is in the Department of Anthropology, University of Utah, 270 S. 1400 E. Stewart 102, Salt Lake City, Utah 84102, USA. e-mail: hawkes@anthro.utah.edu

1. Lahdenperä, M., Lummaa, V., Helle, S., Tremblay, M. & Russell, A. F. *Nature* **428**, 178–181 (2004).
2. Oeppen, J. & Vaupel, J. W. *Science* **296**, 1029–1031 (2002).
3. Nishida, T. *et al. Am. J. Primatol.* **59**, 99–121 (2003).
4. Hill, K. *et al. J. Hum. Evol.* **39**, 1–14 (2001).
5. Kaplan, H. K. *Pop. Dev. Rev.* **20**, 753–791 (1994).
6. Harvey, P. H., Read, A. F. & Promislow, D. E. L. *Oxford Surv. Evol. Biol.* **6**, 13–31 (1989).
7. Ricklefs, R. E. *Am. Nat.* **152**, 24–44 (1998).
8. Charnov, E. L. *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology* (Oxford Univ. Press, 1993).
9. Alvarez, H. P. *Am. J. Phys. Anthropol.* **133**, 435–450 (2000).
10. Goodall, J. *The Chimpanzees of Gombe* (Harvard Univ. Press, Cambridge, MA, 1986).
11. O'Connell, J. F., Hawkes, K. & Blurton Jones, N. G. *J. Hum. Evol.* **36**, 461–485 (1999).
12. Lee, R. D. *Proc. Natl Acad. Sci. USA* **100**, 9637–9642 (2003).
13. Hawkes, K. *Am. J. Hum. Biol.* **15**, 380–400 (2003).
14. Hawkes, K., O'Connell, J. F. & Blurton Jones, N. G. in *Primate Life Histories and Socioecology* (eds Kappeler, P. & Pereira, M.) 204–227 (Univ. Chicago Press, 2003).
15. Doblhammer, G. & Oeppen, J. *Proc. R. Soc. Lond. B* **270**, 1541–1547 (2003).
16. Ricklefs, R. E. & Wikelski, M. *Trends Ecol. Evol.* **17**, 462–468 (2002).
17. Coale, A. J. & Demeny, P. *Regional Model Life Tables and Stable Populations* 2nd edn (Princeton Univ. Press, 1983).

Quantum information

Flight of the qubit

Eugene Polzik

A trapped ion emits a photon. Ion and photon are entangled, so the photon carries away information on the state of the ion. Now realized, this system could become a communication link in a quantum network.

A quantum computer would be capable of performing certain operations, such as factoring large numbers, exponentially faster than its classical counterpart. It could also efficiently model processes that are excessively difficult to model using existing technology. Building such a computer is a formidable task, but spectacular progress has been made in the past few years. It is now widely acknowledged

that one of the most promising systems for quantum computation is an array of ions, trapped and controlled inside an electric field. The initial proposal¹ by Ignacio Cirac and Peter Zoller in 1995 has since been followed up by a train of theoretical and experimental breakthroughs, which last year arrived at the demonstration of elementary quantum logic gates using trapped ions^{2,3}. Now (page 153 of this issue), Blinov

*et al.*⁴ report the first observation of entanglement between a trapped ion and light — a significant step towards building a quantum network.

Entanglement is a quantum correlation between various parts of a system and is required for processing quantum information. The quantum logic gates^{2,3} involving trapped ions were built with short-range entanglement, over only a few micrometres, created by electrical interaction between the ions. Such short-range interaction is not suitable for linking distant nodes of a quantum computer, let alone a large-scale network of such computers. Quantum networks should be linked with light, which is the best long-distance carrier of information, be it classical or quantum.

In a classical computer, bits of information are physically implemented as charges on tiny capacitors, and can take two distinct values, usually denoted 0 and 1. Quantum mechanics, however, allows for a superposition of states (written as $|0\rangle + |1\rangle$). This superposition, called a quantum bit or 'qubit', dramatically enhances computing and communication capabilities.

More specifically, in the work of Blinov *et al.*⁴, a qubit formed by a trapped ion can exist in a superposition of two different orientations of its magnetic momentum, say 'up' and 'down', or $+1$ and -1 . The superposition is then described as $\alpha|-1\rangle + \beta|+1\rangle$, where α and β are coefficients normalized so that the probability of the ion being somewhere is one: $|\alpha|^2 + |\beta|^2 = 1$. The trick by which Blinov *et al.* entangled an ion and a photon was to bring the ion into this superposition of states through the emission of the photon.

According to the conservation of angular momentum, if the ion is created in a spin-down state $|-1\rangle$, the emitted photon is circularly polarized (a property of its electric-field vector) in the right-hand direction — let's call this state of photon polarization $|->$. Similarly, if the ion is created in a spin-up state $|+1\rangle$, the emitted photon has left-hand circular polarization ($|+>$). Most importantly, if the ion ends up in an unknown superposition of spin-up and spin-down states, the emitted photon has a complementary superposition state. This is the essence of entanglement of two qubits, which can be written as a joint state for the ion-photon system, $\alpha|-1\rangle|-> + \beta|+1\rangle|+>$.

Such an entangled state has been generated before, most often between two photons, but also for two ions^{2,3}, two atoms⁵ or an atom and a microwave photon in a cavity⁶. In fact, entanglement between atoms and light has been hinted at in a variety of experiments: for example, in the quantum correlations in light emitted by atomic ensembles^{6,7}; in work on spin squeezing⁸ and the entanglement of atomic ensembles⁹; and in early experiments

on Bell inequalities, in which two photons were emitted by a single atom^{10,11}.

The real breakthrough achieved by Blinov *et al.*⁴ is that, for the first time, entanglement has been observed between a stationary computational qubit (a trapped ion) and a 'flying' communication qubit (an optical photon). The emitted photon can carry a unique piece of information about the state of the ion over a long distance. Another advantage of the system is that trapped ions have exceptionally long lifetimes in entangled states. Although in this experiment the lifetime of demonstrated entanglement did not exceed a microsecond, it can potentially be increased by many orders of magnitude.

Note that the photon was emitted by the ion spontaneously. This means that the photon's state of polarization was random, as was the direction in which it was emitted. Blinov *et al.* excited the trapped ion in such a way that they could count the emitted photons one by one. They then chose to detect photons emitted only with a specific polarization and in only one direction. Although this selected only a small fraction of all photons emitted, when their detector registered a photon with a particular polarization Blinov *et al.* knew that the ion had been left in a well-defined superposition of states. (In principle, the choice of polarization at the photon detector could have been made after the photon was emitted, to demonstrate the nonlocality of the entangled photon-ion system via the violation of Bell inequalities.)

The fact that the generation of the entangled state was conditional on the detection of the photon to some extent limits the

applications of this technique. However, that is not an obstacle to what seems to be the most promising application — the entanglement of two distant ions through the joint detection of two photons emitted by the two ions ('distant' in this context means tens of metres; the distance could be extended to many kilometres by using a different photon wavelength). Under present experimental conditions, the rate for this process would be small (proportional to the square of the already small rate of single-photon detection), but it could be increased by placing the ions in optical cavities. From there, the system could be scaled up by entangling each ion with its neighbour inside the trap, and a chain of such quantum circuits — a 'quantum repeater'¹² — could be constructed. Ultimately, a quantum link over a very long distance could be created. That's some way ahead, but Blinov *et al.*⁴ have taken us a few more steps in this exciting direction. ■

Eugene Polzik is at the Niels Bohr Institute and at the Danish National Research Foundation Center for Quantum Optics (QUANTOP), Blegdamsvej 17, Copenhagen 2100, Denmark.

e-mail: polzik@nbi.dk

1. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
2. Liebfried, D. *et al. Nature* **422**, 412–415 (2003).
3. Schmidt-Kaler, F. *et al. Nature* **422**, 408–411 (2003).
4. Blinov, B. B., Moehring, D. L., Duan, L.-M. & Monroe, C. *Nature* **428**, 153–157 (2004).
5. Rauschenbeutel, A. *et al. Science* **288**, 2024–2028 (2000).
6. Kuzmich, A. *et al. Nature* **423**, 731–734 (2003).
7. van der Wal, C. H. *et al. Science* **301**, 196–200 (2003).
8. Kuzmich, A., Mandel, L. & Bigelow, N. P. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
9. Julsgaard, B., Kozhekin, A. & Polzik, E. S. *Nature* **413**, 400–403 (2001).
10. Freedman, S. J. & Clauser, J. F. *Phys. Rev. Lett.* **28**, 938 (1972).
11. Aspect, A. *et al. Phys. Rev. Lett.* **47**, 460–463 (1981).
12. Briegel, H. J., Dur, W., Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).

Global change

An Earth on fire

Helmut Weissert and Stefano M. Bernasconi

Fifty-five million years ago the Earth suddenly got much hotter. Events are recorded in a 'spike' in the carbon-isotope record, for which a provocative new explanation has been proposed.

In 1997–98, extreme El Niño climatic conditions in the tropical Pacific had severe consequences in Indonesia. A prolonged period of dry weather resulted in drought, and favoured the ensuing forest and peat fires (Fig. 1) that were among the largest of the past century. Massive amounts of carbon were released into the atmosphere, which led to the biggest annual increase in atmospheric CO₂ of the past 50 years¹. Could natural wildfires, developing under more extreme climatic conditions, have had an even larger impact on the carbon cycle in the geological past? Writing in *Paleoceanography*, Kurtz and colleagues² propose that they did.

Kurtz *et al.* argue that 55 million years

ago, during a time of dry climate, peat fires could have burned for thousands of years. They claim that these global wildfires released huge amounts of carbon into the atmosphere, which are recorded in a prominent negative 'spike' in the marine carbon-isotope record. This spike coincides with an episode known as the Palaeocene–Eocene thermal maximum, in which temperatures soared worldwide. This new model contrasts with the now widely accepted view that the negative carbon-isotope spike was produced by a huge burst of methane³, triggered by climate-induced destabilization of methane-bearing gas hydrates beneath the sea floor.

The provocative argument developed by



Figure 1 Indonesia ablaze, 1998. These widespread fires released massive amounts of carbon into the atmosphere — maybe an echo of events 55 million years ago.

Kurtz *et al.* is based on an isotope mass-balance modelling approach, and on the use of two geochemical indicators, the carbon-isotope and sulphur-isotope records. The carbon-isotope record serves as a proxy indicator of the burial history of organic carbon through time, and the sulphur-isotope record provides information on the main storage sites of organic carbon. Organic matter formed in the biosphere is strongly depleted in the heavy isotope ^{13}C . If this organic carbon were buried at increased rates in marine or terrestrial deposits, ^{13}C would have become enriched in the marine carbon reservoir. Such changes in the oceanic carbon pool would be recorded in the carbon-isotope compositions of biologically generated calcium carbonate (CaCO_3) precipitated from sea water, and this isotope signature would be fossilized in marine limestone.

Similarly, the sulphur-isotope signature in marine barium sulphate (BaSO_4) provides a record of the oceanic sulphate pool. At a time of high burial rates of marine organic carbon, the marine sulphate pool should shift to isotopically heavier values due to the increased activity of sulphate-reducing bacteria. Sulphate reduction would lead to an increased burial of ^{34}S -depleted sulphur in the form of pyrite (FeS_2). If the sulphur- and carbon-isotope curves are decoupled, the implication is that organic carbon was preferentially buried on continents, with only limited burial of biologically generated sulphur because of the limited supply of sulphate in freshwater environments.

Kurtz and colleagues² base their mass-balance models on the available sulphur-isotope⁴ and carbon-isotope⁵ records. They first trace carbon burial through the Palaeocene, between about 65 million and 55 million years ago, when the carbon-isotope curve is

marked by the Palaeocene carbon-isotope maximum, a time of elevated values. Like earlier authors, Kurtz *et al.* take these data as evidence for increased burial of organic carbon. Yet marine Palaeocene sediments are not rich in organic matter. So, many investigators have proposed that the Palaeocene was a time of widespread organic-matter sedimentation on continents. Kurtz *et al.* test this hypothesis by using the sulphur-isotope record as an additional source of information. The lack of change in sulphur-isotope values indicates a decoupling of the sulphur and carbon cycles during the time of increased burial of organic carbon, supporting the idea that organic carbon was stored in peatlands.

Kurtz *et al.* integrate these findings into their new scheme of events. They argue that under increasingly dry conditions, and at the time of decreasing peat formation, wildfires may have raged through peatlands for up to thousands of years. Terrestrial organic carbon would have been rapidly oxidized and isotopically light carbon dioxide would have been transferred to the atmosphere and oceans, causing a sudden warming. The negative spike in the carbon isotope curve would then have originated from terrestrial organic carbon.

Can this new hypothesis be further tested? Available wildfire records are poor and only future investigations will show whether the Palaeocene–Eocene thermal maximum was characterized by increased charcoal deposition, or whether there are traces of fire in the remaining Palaeocene coal deposits. Perhaps it was not only surface wildfires, but also extensive subsurface burning of organic matter derived from abundant lake sediments, as recently documented from Mali⁶, that contributed to the perturbation of the



100 YEARS AGO

All the experimental evidence so far obtained is now in agreement with the view that the γ rays are an extremely penetrating type of Röntgen rays which have their source in the atom of the radio-active substance at the moment of the expulsion of the β or cathodic particle. For example, I have found that the γ rays from radium always accompany the β rays, and are always proportional in amount to them. In radium the β and γ rays appear only in the third change occurring in the radio-active matter which causes “excited activity,” *i.e.* in the fourth of the chain of radio-active products which result from the disintegration of the radium atom... The γ rays arise from the disintegrated atom, and are not secondary rays set up by the bombardment of the radium as a whole by the β rays.

E. Rutherford

From *Nature* 10 March 1904.

50 YEARS AGO

During the past year, evidence has gradually accumulated for the existence of a new type of nuclear excitation which appears to be due to the presence, within a nucleus, of a neutral hyperon, now designated Λ^0 . It is believed that this particle can be bound to the other nucleons of a nucleus to form a relatively stable structure as measured on a nuclear time-scale, and that the eventual disintegration of such a nucleus is due to the decay of the Λ^0 -particle. The existence of the Λ^0 -particle, which until recently was referred to as the ‘heavy neutral V -particle’, was established by experiments on the cosmic radiation with Wilson chambers. It was shown to be the more common of the neutral particles which produce the ‘neutral V -events’ discovered by Rochester and Butler in 1947... These considerations suggest that the Λ^0 -particle is an excited nucleon in a different sense from that suggested by familiar analogies. We are entering a new field where basically new concepts remain to be established; but it seems reasonable to conjecture that the nucleon is transformed into an excited nucleon as a result of changes in its internal constitution. If so, we are beginning to make a new penetration into what Maxwell called “the strange strata of the material world”, a penetration into the world of the nucleon. It seems that matter is inexhaustible.

C. F. Powell

From *Nature* 13 March 1954.

carbon cycle. If traces of global fires are not found, rapid release of methane will remain the most likely cause of the negative carbon-isotope spike⁷. But Kurtz *et al.* have made us aware that an integrated approach, one that combines studies of the carbon cycle with research into the evolution of sulphur cycle, could provide new insights into Earth's environmental history. ■

Helmut Weissert and Stefano M. Bernasconi are in

the Department of Earth Sciences, ETH, CH-8092 Zurich, Switzerland.

e-mail: weissert@erdw.ethz.ch

1. Page, S. E. *et al.* *Nature* **420**, 61–65 (2002).
2. Kurtz, A. C., Kump, L. R., Arthur, M. A., Zachos, J. C. & Paytan, A. *Paleoceanography* **18**, 1090 (2003).
3. Dickens, G. R. *et al.* *Paleoceanography* **10**, 965–971 (1995).
4. Paytan, A. *et al.* *Science* **282**, 1459–1462 (1998).
5. Oberhänsli, H. & Hsu, K. *Geodyn. Ser.* **15**, 85–100 (AGU, Washington DC, 1986).
6. Svensen, H. *et al.* *Geology* **31**, 581–584 (2003).
7. Weissert, H. *Nature* **406**, 356–357 (2000).

Astrophysics

The inconstant constant?

Lennox L. Cowie and Antoinette Songaila

Has the value of the fine-structure constant changed over the history of the Universe? An earlier analysis of radiation from distant quasars suggested the answer is yes — a new analysis says no.

The physical constants might not be so constant. If any variation in their values were measured, it could give us profound constraints on a long-sought quantized theory of gravity. So it is no surprise then that a claim^{1,2} to have measured a variation over time in the value of the fine-structure constant, α , has led to a spate of papers incorporating this result into a wide range of theories. Now Chand *et al.*, in a paper³ to appear in *Astronomy and Astrophysics*, have used the same technique of measuring radiation from distant quasars but reach the opposite conclusion — there is no significant variation in α . What should we believe?

The fine-structure constant was originally uncovered in studies of the closely spaced spectral lines of atoms such as hydrogen and helium. This 'fine structure' reflects the quantization of electron energies within the atom. The constant is defined as the product $2\pi e^2/hc$ (where e is the charge of the electron, h is Planck's constant and c is the speed of light) but is perhaps more familiar in its numerical form: $\alpha = 1/137$.

Curiously enough, the most stringent limits on the variation of α come not from laboratory or astronomical measurements but from a natural nuclear reactor in Africa. About 1.8 billion years ago, in what is now the Oklo mine in Gabon, a spontaneous chain reaction began, involving the fission of the uranium isotope ²³⁵U. The capture of thermal neutrons from the fission process by other elements can be related to α . The best recent analysis of data from Oklo shows that any fractional change in the fine-structure constant ($\Delta\alpha/\alpha$) has been less than 10^{-7} over nearly two billion years⁴.

Although less sensitive, cosmological measurements can nevertheless probe much longer periods, encompassing much of the 14-billion-year life of the Universe. Gas in and between galaxies that lie along the line of

sight from the Earth to quasars scatters the light from these enormously distant sources. Atoms and molecules in the gas absorb certain wavelengths, imprinting absorption lines in the radiation spectra of these objects. Often the lines of many elements, in many ionization states, might be seen from a particular patch of gas. The wavelengths at which the lines occur depend on the distance of the absorbing gas from the observer, because the radiation becomes 'redshifted' to longer wavelengths, owing to the expansion of the Universe, as it travels from its source.

If there had been any variation in the fine-structure constant over the billions of years of the light's journey, that would have affected the energy levels in the atoms and would therefore have shifted the wavelengths of the absorption lines. We cannot measure absolute shifts in these wavelengths, because we have no way of independently knowing the distance to the source and hence the redshift that the radiation has undergone. But we can measure relative shifts of the wavelengths from all the absorption lines seen for a particular system. Absorption lines have been detected for quasars so distant that the radiation we see from them corresponds to a time when the Universe was only 6% of its present age^{5,6}.

The simplest possible measurement is to look at lines that are all associated with the same ion — this is the basis of the so-called alkaline doublet (AD) method. Here, the separation of the two lines is proportional to the square of the fine-structure constant, so a change in α produces a change in the separation that is proportional to α . However, the limiting sensitivity of this method, even averaging over a large number of systems, is only about 10^{-5} , and no evidence of variation has ever been found at this level^{7,8}.

John K. Webb and colleagues^{9,10} recognized that a much more sensitive check on α could be made by combining measurements

from different atomic species. This technique, which they call the 'many multiplet', or MM, method, can reach sensitivities to $\Delta\alpha/\alpha$ of 10^{-6} — a tenfold improvement over the AD method. This improvement comes at a cost, however. When different ions are compared, they might lie at slightly different positions, and therefore different redshifts, in the absorbing gas; we have to assume that this type of effect averages out when measuring many independent quasar absorption-line systems.

Even more subtle systematic effects might creep in. The fractional mix of the isotopes of a particular element might change with time, causing small systematic shifts in the lines from that element that could be misinterpreted as a change in α . Webb *et al.* considered these possibilities in a series of papers but concluded that, despite this, there really was evidence¹ of a change in α . Their most recent measurement² gives a significant fractional change of $(-0.57 \pm 0.11) \times 10^{-5}$ over the past 8 billion years. This would require that the evolution of α has not been linear to be consistent with the more stringent constraint from the Oklo data.

The physics community has been intrigued by this result, but the astronomical community has been somewhat sceptical. To some extent this is a cultural issue. Astronomers are generally rather wary of complicated physics except when it fits into traditional conceptual models. In this regard, the recent measurements¹¹ of the cosmological constant fall squarely within the astronomical tradition of measuring cosmological geometry and therefore seem quite natural; the variation in the fine-structure constant, however, is regarded as something of an oddball. By contrast, most physicists were probably more surprised that the cosmological constant was not zero than by the claimed variation in the fine-structure constant.

But astronomers' scepticism is also based on a knowledge of how complex the quasar absorption line systems are and how difficult they are to interpret. Astronomers generally rely on multiple lines of evidence to become comfortable with a result. The MM-method claim, however, is based on one set of measurements of a rather heterogeneous sample by a single group.

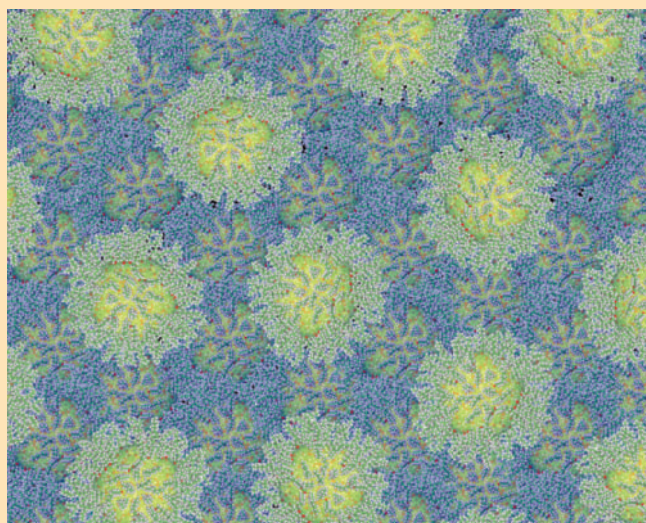
Now it seems that that scepticism might be justified. Chand and colleagues' new measurement³ also uses the MM method, but with a sample of simpler absorption-line systems observed with a higher signal-to-noise ratio. Chand *et al.* have a null result — $(-0.06 \pm 0.06) \times 10^{-5}$ for the fractional change in α , which is inconsistent with the previous measurement^{1,2}. It is not easy to understand how such a different result arises, which might be another reason for scepticism. Possibly there are systematic errors in one or other of the sets of measurements, or

Condensed-matter physics

Supramolecular twelve-a-side

Quasicrystals are freaks of the crystal world. Strictly speaking, they are not crystalline: although the five-fold symmetry some of them exhibit in two dimensions might be acceptable for, say, a starfish, it is not permitted in a perfect crystal. For example, a floor can be tiled using only triangles, or squares or hexagons; tiles with five-fold, eight-fold or twelve-fold symmetry, however, will leave gaping holes if flat, or a puckered surface if forced to fit.

But in this issue, Xiangbing Zeng *et al.* report their discovery of a highly ordered molecular quasicrystal that has the 'forbidden' twelve-fold, or dodecagonal, symmetry (*Nature* **428**, 157–160; 2004). All but one of the quasicrystals known so far are metallic alloys (the notable exception is a liquid-crystal film with



non-crystallographic helical symmetry). Instead of metallic atoms, the new quasicrystal is formed from organic dendritic (tree-like) structures that, Zeng *et al.*

propose, have self-assembled into supramolecular spheres, or micelles.

This organic superstructure (shown here in simulation) was unexpected. Its stability may well

have broader implications for cellular organization in tissues, for instance, or for a problem proposed by Lord Kelvin in 1887: in a foam, what shape of individual, equal-volume bubbles would yield the minimum surface area for the foam? Kelvin suggested a warped, 14-sided structure (a truncated octahedron) but was not able to prove it.

More than a hundred years later, Robert Phelan and Denis Weaire showed that a lower overall surface area would result if a foam contained two kinds of bubbles, one with 14 sides and the other with 12 (*Phil. Mag. Lett.* **69**, 107–110; 1994). With their dodecagonal quasicrystal in mind, Zeng and colleagues suggest that this classic packing problem might now have a solution based on quasicrystalline structure.

May Chiao

possibly the errors are being underestimated by both groups. Whatever the explanation, for the moment the evidence for any variation in the fine-structure constant looks very weak indeed.

Lennox L. Cowie and Antoinette Songaila are at the Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.

e-mail: cowie@mach.ifa.hawaii.edu

1. Webb, J. K. *et al.* *Phys. Rev. Lett.* **87**, 091301 (2001).

2. Murphy, M. T., Webb, J. K. & Flambaum, V. V. *Mon. Not. R. Astron. Soc.* **345**, 609–638 (2003).

3. Chand, H., Srianand, R., Petitjean, P. & Aracil, B. *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/astro-ph/0402177> (2004).

4. Damour, T. & Dyson, F. *Nucl. Phys. B* **480**, 37–54 (1996).

5. White, R. L., Becker, R. H., Fan, X. & Strauss, M. A. *Astron. J.* **126**, 1–14 (2003).

6. Songaila, A. *Astron. J.* (in the press); preprint at <http://arxiv.org/abs/astro-ph/0402347> (2004).

7. Cowie, L. L. & Songaila, A. *Astrophys. J.* **453**, 596–598 (1995).

8. Bahcall, J. N., Steinhardt, C. L. & Schlegel, D. *Astrophys. J.* **600**, 520–543 (2004).

9. Webb, J. K., Flambaum, V. V., Churchill, C. W., Drinkwater, M. J. & Barrow, J. D. *Phys. Rev. Lett.* **82**, 884–887 (1999).

10. Dzuba, V. A., Flambaum, V. V. & Webb, J. K. *Phys. Rev. Lett.* **82**, 888–891 (1999).

11. Spergel, D. N. *et al.* *Astrophys. J. Suppl. Ser.* **148**, 175–194 (2003).

of 30 years produce an increasing fraction of defective oocytes (Fig. 1, overleaf). Consequently, researchers have searched for factors that allow functional oocytes to be stored for 30 years but not for 40 years, with little success. An article by Johnson *et al.*⁴ on page 145 of this issue now demonstrates that, contrary to long-held views, female mice contain a population of germline stem cells that are required to maintain overall follicle numbers during adult life. This important finding seems destined to greatly enhance our understanding of mammalian oogenesis and of its precipitous decline during adulthood.

Stem cells within the ovaries of mice might have been missed in earlier studies because, like other stem cells, they are expected to be rare and cannot be recognized definitively just on the basis of morphological criteria or by the use of known molecular markers. So Johnson *et al.*⁴ instead carefully measured follicle numbers and their rate of loss after birth. As in previous studies, juvenile mice from several different mouse strains had 2,500–5,000 healthy follicles. But the number of dying follicles increased markedly, beginning about 30 days after birth, to as many as 1,200 per ovary and remained high until the mice were at least 4 months old. Dying follicles degenerate within a few days, which means that the follicular loss rate must be high. As the total number of healthy follicles decreased relatively little (if at all) despite this rapid follicle turnover, the authors concluded that new

Stem cells

More like a man

Allan C. Spradling

Most female mammals experience a reproductive decline with increased age, previously attributed to the instability of ageing oocytes. But could it be due to a previously unrecognized stem-cell well drying up?

Why can't a woman be more like a man? Male mammals generally can reproduce throughout most of their adult lives by continuously generating sperm precursors from germline stem cells maintained within the testis. In contrast, most mammalian females show a reproductive decline as they get older. Females of the few species that remain fertile throughout life, such as the well-studied fruitfly *Drosophila melanogaster*, contain

germline stem cells like those of males, and use them in the constant replenishment of oocyte precursors¹.

Fifty years ago, mammalian ovaries were also thought by many to contain stem cells in an outer layer², but since then the stability of a fixed supply of follicles (growing oocytes encircled by support cells) produced during fetal life has been believed to determine the reproductive lifespan³. Women are fertile only when relatively young, and after the age

follicles must be being made somewhere in young mouse ovaries. They proposed an active population of germline stem cells as the most likely source.

Four lines of evidence from the new study⁴ strongly corroborate the existence of germline stem cells and the resultant ongoing follicle production in the ovaries of postnatal mice. First, Johnson *et al.* found a population of 63 ± 8 germ cells near the surface of the ovary that were outside follicles and that had the general characteristics of germline stem cells. They expressed a conserved germ-cell marker known as Vasa and were actively cycling — that is, they produced new cells.

Second, the authors found germ cells that had not yet developed into follicles in the ovary. This would be expected if the progeny of the germline stem cells were developing into new oocytes; indeed, Johnson *et al.*⁴ identified germ cells that expressed several markers of this process. The overall levels of these marker proteins in the ovaries amounted to 6–25% of the levels that occur in adult testes, which contain large numbers of cycling germline stem cells and their maturing progeny (which express these same proteins).

Third, that the progeny of germ cells went on to produce follicles was shown by grafting a small part of a fluorescently labelled ovary into an unmarked host ovary. Follicles that comprised fluorescently labelled oocytes surrounded by unmarked follicle cells could subsequently be found. Germ cells had therefore moved out of the graft and formed a follicle within the host ovary, something that would be expected only if follicle formation was ongoing.

Last, treating mice with the toxic drug busulfan, which selectively kills male germline stem cells, rapidly depleted the pool of young follicles. The rapid time frame in which it did so suggested that female germline stem cells in mice contribute an average of 77 new follicles per ovary per day. The group's findings⁴ all point to the existence of germline stem cells and a requirement for them in maintaining follicle numbers throughout the reproductive lifetime of female mice.

The finding raises many interesting and important questions. Determining the exact number and location of these functional germline stem cells will require tagging germ cells and showing that their descendants form follicles and, ultimately, mature oocytes. Such lineage-tracing experiments will also address whether the progeny of mouse germline stem cells differentiate directly into oocytes or whether they first increase their numbers by forming interconnected 'germ-cell cysts'. These structures are present during fetal stages⁵ and form after germline stem-cell division in most organisms.

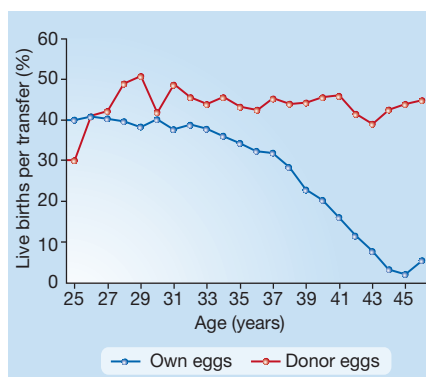


Figure 1 Falling fertility. The graph shows the decreasing success rates (live births) with increasing age for *in vitro* fertilization using embryos derived from a woman's own eggs compared with using eggs from a young donor.

Another issue will be to assess the relative use of follicles generated during fetal life compared with those produced by adult stem cells. Do the follicles produced before birth fail to survive to reproductive maturity, so that female fertility actually depends on the presence of young follicles produced recently from germline stem cells? And does the loss of these stem cells soon lead to follicle ageing, depletion and reproductive decline?

Now that germline stem cells in the mouse have been discovered, determining the cellular and molecular mechanisms that maintain them as such will be the mission of many a scientist. Previous work on *Drosophila* germline stem cells might usefully guide this quest. These stem cells reside in a niche where certain conditions are met to ensure that they continue to survive and function.

Planetary science

Secrets of the deep

Jonathan Aurnou

The magnetic fields of Uranus and Neptune are markedly different from those of other planets in the Solar System. Can this be attributed to structural differences deep inside the planets?

Several planets in the Solar System — including Jupiter and its moon Ganymede, Saturn and possibly Mercury¹ — have a magnetic field that is similar to Earth's. The magnetic fields resemble that of a bar magnet, with the alignment of north and south poles oriented close to the rotation axis of the planet. But data from NASA's Voyager 2 probe have shown that the magnetic fields of Uranus and Neptune are different from those of other planets. Their fields are effectively tipped over: instead of aligning along the rotational axis, the north–south axis of the field lies midway or closer to the equatorial

Germline stem cells have several requirements: contact with a specific body (non-gamete) cell type; a particular signal, a member of the 'bone morphogenetic protein' family; and the expression of a small number of special regulatory genes^{1,6}. Many of these mechanisms have probably been conserved between species. All these issues can now be addressed in mice by using existing technology.

Last, but far from least, the question on everyone's lips will be whether there are germline stem cells in the human ovary. Germline stem cells in humans might easily have been missed for the same reasons that they escaped detection in mice for so long. Indeed, the work of Johnson *et al.*⁴ raises the strong possibility that the reproductive decline seen in female thirtysomethings is due to the depletion of germline stem cells coupled with a high follicular age-dependent incidence of defects occurring during reductive divisions when oocytes mature⁷.

Allan C. Spradling is in the Howard Hughes Medical Institute Laboratory, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, Maryland 21210, USA.

e-mail: spradling@ciwemb.edu

1. Spradling, A., Drummond-Barbosa, D. & Kai, T. *Nature* **414**, 12–18 (2001).
2. Raven, C. P. *Oogenesis: The Storage of Developmental Information* (Pergamon, New York, 1961).
3. Tilly, J. L. *Nature Rev. Mol. Cell Biol.* **2**, 838–848 (2001).
4. Johnson, J., Canning, J., Kaneko, T., Pru, J. K. & Tilly, J. L. *Nature* **428**, 145–150 (2004).
5. Pepling, M. E. & Spradling, A. C. *Development* **125**, 3323–3328 (1998).
6. Song, X., Zhu, C. H., Doan, C. & Xie, T. *Science* **296**, 1855–1857 (2002).
7. Pellestor, F., Andréo, B., Arnal, F., Humeau, C. & Demaille, J. *Hum. Genet.* **112**, 195–203 (2003).

plane. These unusual magnetic fields have been difficult to model. But on page 151 of this issue, Stanley and Bloxham's simulations² show that, by altering the description of the internal structure of the planets, complex magnetic fields can be generated that are similar in structure to those of Uranus and Neptune.

All models of the generation of planetary magnetic fields include the same essential ingredients. There must be a region of electrically conducting fluid and an energy source to drive the motion of that fluid. A model of Earth's field, for instance, simulates its iron-rich molten outer core (the

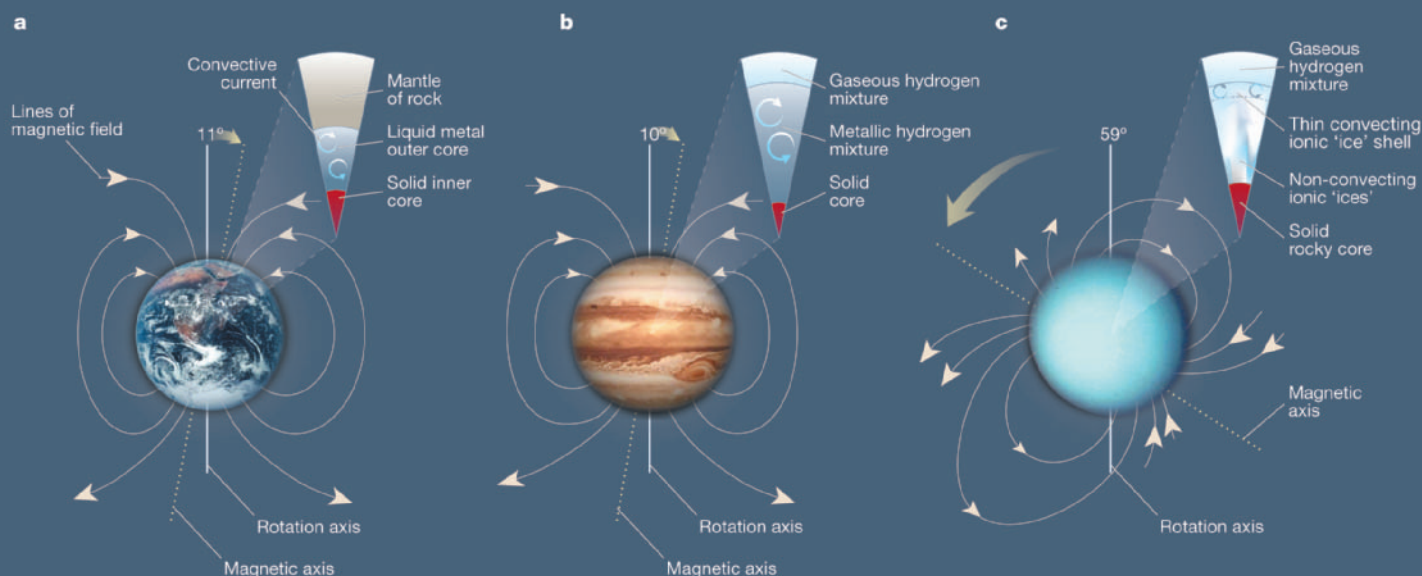


Figure 1 Magnetic fields and interior structures of the planets. **a**, Earth has a predominantly dipolar magnetic field, like that of a bar magnet. Magnetic north and the North Pole do not quite coincide: the magnetic axis is tilted by 11° relative to the axis of rotation of the planet. The field is generated by the dynamo action of convective currents in the molten outer core, around a solid conducting inner core. **b**, The gas giant Jupiter also has a strongly dipolar field, whose axis is aligned at 10° to the planet's rotational axis. In Jupiter's interior, convection occurs in a thick shell of metallic (dissociated) hydrogen that surrounds the

planet's small, rocky core. **c**, Uranus, in contrast, has comparable quadrupole and dipole components in its magnetic field. Strikingly, the magnetic-field axis is oriented at 59° to the planet's axis of rotation. Using a revised description of the planet's interior — with a much thinner convecting shell around a non-convecting but fluid inner region — Stanley and Bloxham² have devised a dynamo model that generates a magnetic field that is similar in structure to Uranus' (and Neptune's) anomalous field. (Rotation axes are artificially aligned vertically for ease of comparison.)

conducting fluid) as well as the planetary cooling (or radioactive heating) that drives convective currents in the planet's interior. Another necessary element is the planetary rotation that organizes the fluid motions: well-organized fluid motions can generate a large-scale magnetic field, but disorganized fluid motions produce fields that tend, on average, to cancel out. Only when all these conditions are met, in a model or a planet, is it possible for the moving fluid to create a dynamo that generates a planetary magnetic field³.

For Earth-like planets, convective motions are typically modelled in a thick rotating shell of electrically conducting fluid. That shell surrounds a relatively small, electrically conducting, solid inner core (Fig. 1a). The result is a dipolar, bar-magnet style magnetic field that is almost aligned with the rotation axis. A similar set-up holds for the gas giants Jupiter and Saturn (Fig. 1b): a very small rocky core is surrounded by a thick layer of convecting metallic hydrogen. (Hydrogen under very high pressure dissociates into free-moving protons and electrons, and is hence called 'metallic'.)

But such models are largely unable to capture the complexities of the magnetic fields of Uranus and Neptune. The measurements made by Voyager 2 around these planets revealed that their fields are not predominantly dipolar (like a bar magnet),

but also have a quadrupole component (as though produced by a combination of two bar magnets, with two north and two south poles). In addition, the main axis of their fields lies far from the rotation axis. On Uranus, the main magnetic-field axis is tipped over from the rotation axis by about 59° in latitude; on Neptune, the main axis is tipped over by about 47° .

Stanley and Bloxham² have constructed a model that simulates these characteristic differences. Instead of a thick convecting shell and a solid inner region, they consider Uranus and Neptune to have a comparatively thin outermost region of convecting ionized fluid surrounding a non-convecting ionized inner-fluid 'ocean'. This modified structure follows the suggestion of earlier work^{4,5}, based on anomalously low planetary heat-flow measurements, that deep convective fluid motions on Uranus and Neptune might be occurring in only a relatively thin shell of the planetary interior (Fig. 1c).

Stanley and Bloxham's model generates fields similar to those observed on Uranus and Neptune. Indeed, it seems that a relatively simple alteration in the structure of the convecting region can lead to a fundamental change in the type of magnetic field that a dynamo model produces. Why this fundamental change in magnetic-field structure should occur remains unclear. Stanley and Bloxham propose that the less constraining

effects of the fluid inner region allow more complex magnetic-field morphologies to develop. However, this argument does not explain how the complex magnetic fields are produced and maintained.

It also remains to be seen how robust this complex field behaviour is. Stanley and Bloxham present a single-case result in their study. Are complex fields produced in all cases with this internal structure, or only, for instance, in those that have convective motions of a particular strength? These questions aside, their model clearly and impressively demonstrates that the same fundamental process — convection in a rotating spherical shell of electrically conducting fluid — can explain the basic structure of all the dynamo-generated planetary magnetic fields presently observed in the Solar System. ■

Jonathan Aurnou is in the Department of Earth and Space Science, University of California, Los Angeles, 595 Charles Young Drive East, Los Angeles, California 90095-1567, USA.
e-mail: aurnou@ucla.edu

1. Solomon, S. C. *et al.* *Planet. Space. Sci.* **49**, 1445–1465 (2001).
2. Stanley, S. & Bloxham, J. *Nature* **428**, 151–153 (2004).
3. Busse, F. H. *Phys. Fluids* **14**, 1301–1314 (2002).
4. Podolak, M., Hubbard, W. B. & Stevenson, D. J. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 29–61 (Univ. Arizona Press, 1991).
5. Hubbard, W. B., Podolak, M. & Stevenson, D. J. in *Neptune and Triton* (ed. Cruikshank, D. P.) 109–138 (Univ. Arizona Press, 1995).

Animal behaviour

Birth control by royal pheromones in ants

Proc. Natl Acad. Sci. USA **101**, 2945–2950 (2004)

In an ant colony with a fertile queen, worker ants don't expend their energy in breeding — they save it for colony upkeep. The queen somehow conveys this message to worker ants despite not coming into direct contact with them. Annett Endler *et al.* wondered how this signal is spread and found the answer to be a pheromone on the queen's eggs.

The researchers set up colonies of the ant species *Camponotus floridanus* without their queen. Without the queen's eggs, the workers started producing their own eggs. But when the queen's eggs were present, they did not produce their own eggs. Adding both workers' eggs and queen's eggs to a colony caused the worker ants to destroy one another's eggs but leave the queen's eggs intact, hinting that it was the eggs that carried the signal.

But what is so distinct about the queen's eggs? Endler *et al.* identified certain hydrocarbons on the surface of the queen's body and swapped these for different hydrocarbons from the workers' eggs. The workers destroyed significantly fewer eggs when they were disguised in this manner, so Endler *et al.* conclude that the queen's unique hydrocarbon signature marks her omnipotence.

Laura Nelson

Gravity

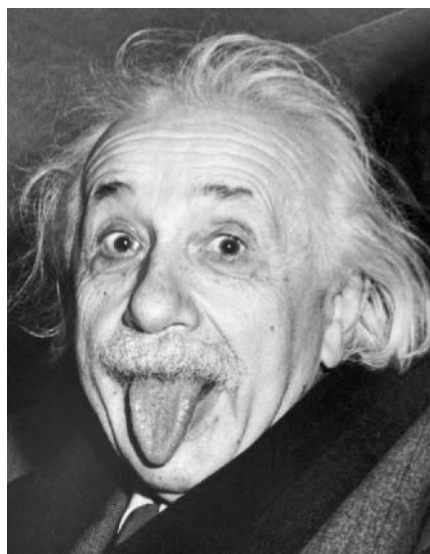
Room for the fat graviton?

Phys. Rev. D **69**, 044014 (2004)

Like Einstein before them, physicists are still rather bothered by the cosmological constant. This seemingly arbitrary add-on to general relativity allows for the expansion of the Universe, and data have confirmed that its value is not zero.

Surely this troublesome number should arise in a more natural way in a more complete theory, string theory perhaps? Clues might lie, not at cosmological distances, but at the micrometre scale. So says Raman Sundrum, who has re-examined the proposal that such a complete theory might include a 'fat graviton': if so, he argues, we could be on the verge of seeing a breakdown in Newton's law of gravity.

All fundamental particles are considered to be point-like. But if the graviton (the as-yet-unseen mediator of the gravitational force) has some finite size, according to Sundrum's estimates its effect might be noticed as a distortion of Newton's law at scales of around 20 micrometres. Experimenters have so far checked out gravity down to around the 100-micrometre mark, and Sundrum



Einstein hated the cosmological constant: could the 'fat graviton' be part of a theory that better justifies its existence?

urges them to continue "to vigorously hunt" for signs of the fat graviton. Even if the graviton turns out to be somewhat slimmer, closing this door to solving the problem of the cosmological constant will be helpful.

Alison Wright

Photonics

Light arrested in forest

Phys. Rev. Lett. **92**, 083901 (2004)

A new way to trap a light beam could prove invaluable for optical computing and quantum information processing. Mehmet Fatih Yanik and Shanhui Fan show how light can be stopped and stored using a lattice of microscopic semiconductor pillars.

In effect, a light pulse passing through this forest of pillars gets stuck by bouncing from pillar to post, until released again if the refractive index of the pillars is altered. Yanik and Fan show that light propagating along a row of optical cavities, made by slimming down individual pillars in the lattice, can be shunted into cavities to either side of the row and held there for a controlled delay time. The stored pulse emerges unscathed — there's no distortion of the signal.

So far, the researchers have only demonstrated their approach using computer simulations. Such semiconductor structures are easy enough to build, but controlling their refractive index might be trickier. 'Stopping' light inside a solid has been reported before (*Phys. Rev. Lett.* **88**, 023602; 2002), but that involved using a laser to tune the electronic states of atoms in a large crystal. Using a microscale pillar array should make the process more compatible with silicon-chip technology.

Philip Ball

Neurodegeneration

Model motor neuron death

Neuron **41**, 687–699 (2004)

Researchers have created a mouse model for a rare human disease called spinal and bulbar muscular atrophy, which causes motor neuron degeneration and muscle wasting. In this disease, an altered gene drives the production of a faulty protein, an androgen receptor, that contains an unusually long run of the amino acid glutamine.

Bryce L. Sopher *et al.* have created mice that have a human androgen receptor containing 100 glutamines. The animals develop progressive limb weakness around mid-adulthood as their motor neurons began to degenerate. These symptoms are strikingly similar to those seen in human spinal and bulbar muscular atrophy.

Not only do the mice provide a useful model of the disease, they also shed light on the mechanisms of motor neuron death. The spinal cords of these mice showed reduced levels of vascular endothelial growth factor, a molecule known to aid motor neuron survival, and when the researchers treated degenerating motor neurons in culture with this growth factor, the cells recovered. So vascular endothelial growth factor may play an essential role in motor neuron degeneration.

Helen R. Pilcher

Environmental science

Sulphur from soil

Atmos. Environ. **38**, 1473–1480 (2004)

Coastal soils can be significant sources of atmospheric sulphur dioxide (SO₂), according to Bennett C. T. Macdonald and colleagues. Soils have long been recognized as sinks for sulphur. But the authors believe that this is the first direct measurement of SO₂ emissions from soils containing sulphide minerals, mostly in the form of pyrites (FeS₂).

Coastal lowland areas often contain sulphidic soils deposited since the last major ice age. When the wet soils are drained, the sulphides oxidize and produce a mixture of sulphur species — some of which apparently end up as SO₂.

Macdonald *et al.* estimate that, globally, these soils may emit about three million tonnes of sulphur each year. This is roughly the amount produced by ships, as estimated by others, and is thought to cause elevated SO₂ levels in coastal areas. Much of this coastal SO₂ may come from soils rather than ships, say Macdonald and colleagues. But humans are still implicated in the emissions — the soils release more SO₂ when they are disturbed or drained as part of land reclamation.

Mark Peplow

Live birth after ovarian tissue transplant

Fresh pieces of monkey ovary remain fully functional even when moved to a new site.

Radiation and high-dose chemotherapy may render women with cancer prematurely sterile, a side-effect that would be avoided if ovarian tissue that had been removed before treatment could be made to function afterwards. Live offspring have been produced from transplanted ovarian tissue in mice¹ and sheep² but not in monkeys³ or humans^{4,5}, although sex steroid hormones are still secreted. Here we describe the successful transplantation of fresh ovarian tissue to a different site in a monkey, which has led to the birth of a healthy female after oocyte production, fertilization and transfer to a surrogate mother. The ectopically grafted tissue functions without surgical connection to major blood vessels and sets the stage for the transplantation of cryopreserved ovarian tissue in humans.

Ovaries were removed laparoscopically from seven anaesthetized rhesus monkeys, *Macaca mulatta*. The ovarian cortex was cut into small pieces ($1 \times 3 \times 4$ mm; $n=219$) in Leibovitz medium at 4°C and transplanted immediately to the animal of origin in subcutaneous pockets or flaps next to muscle or kidney. Four monkeys had transplants to both the arm and abdomen (number of sites, 23–54); two received transplants to the kidney and abdomen (using 18–42 sites); and one received transplants to the arm only (at 26 sites). (See supplementary information for methods.)

Serum oestradiol (E2), progesterone (P4) and follicle-stimulating hormone were measured in samples of the animals' peripheral blood by radioimmunoassay⁶. We monitored animals for menses and follicles by using ultrasound. When 4-mm follicles developed, oocytes were collected by follicle excision 26 to 30 hours after injecting 1,000 international units (IU) of human chorionic gonadotropin. Mature oocytes were fertilized by intracytoplasmic sperm injection⁷



Figure 2 Infant monkey resulting from the transfer to a surrogate mother of two morula-stage embryos. The embryos were generated after fertilization by intracytoplasmic sperm injection of metaphase II-stage oocytes collected from transplanted grafts of ovarian cortex. The infant is known as BRENDA (for bilateral oophorectomy, resumption of endocrine function, and abdominal follicle pregnancy).

after electro-ejaculation. Fertilization was confirmed by the observation of two pronuclei between 10 and 16 hours after sperm injection, with subsequent cleavage of nucleated blastomeres.

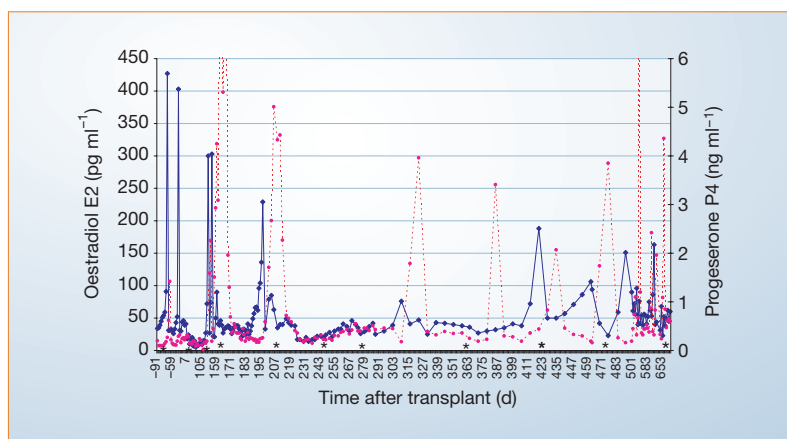
All monkeys had plasma concentrations of E2 over 50 pg ml^{-1} within 70 to 150 days of transplantation. Oestradiol and progesterone values were higher in the local venous drainage of an arm transplant than in systemic venous blood, indicating that the grafted ovarian tissue was functional (see supplementary information). One monkey with renal and abdominal grafts

showed a repeated rise in P4 of more than 3 ng ml^{-1} roughly every 60 days, which is longer than the normal 28-day cycle (Fig. 1). Follicle-stimulating hormone rose in this animal to 10.5 mIU ml^{-1} 84 days after transplantation, but declined to 2.79 mIU ml^{-1} by day 169, indicating that oestrogen production was adequate.

Several animals developed multiple follicles without exogenous gonadotropin stimulation; abdominal subcutaneous grafts showed the best follicular development (50%; see supplementary information). Attempts to retrieve ovarian follicles from the kidney capsule were unsuccessful as this site is difficult to access surgically, and needle aspiration of other follicles failed to yield oocytes. We therefore excised follicles ($n=23$) from four monkeys that had been treated with human chorionic gonadotropin, obtaining 16 oocytes.

Eight of these excised follicles were mature; we fertilized six by intracytoplasmic sperm injection and they cleaved *in vitro*. One 5-cell and one 8-cell embryo were transferred laparoscopically to the ovi-

Figure 1 Systemic oestradiol (E2; blue) and progesterone (P4; pink) levels over 660 days in a monkey that had had ovarian tissue transplanted to the kidney and abdomen. Reproductive-hormone production was cyclic, with the first progesterone peak occurring 166 days after transplantation and with eight subsequent P4 peaks of more than 2 ng ml^{-1} occurring on average every 62 days. Vaginal bleeding, at time points marked with an asterisk, occurred ten times after transplantation, and correlated with falling oestrogen levels.



ducts of two recipient monkeys. Two morula-stage embryos were transferred laparoscopically to a third monkey. A normal singleton gestation resulted from the transfer of the morulas, ending in the birth last year of a healthy female weighing 500 g (Fig. 2).

To our knowledge, this is the first successful fertilization and pregnancy to follow egg collection from fresh transplanted ovarian tissue in a primate. Although the kidney supported endocrine function in the transplanted tissue, the arm and abdomen transplantation sites provided easier access for retrieving oocytes.

The procedure described here has the potential to rescue fertility in cancer survivors. However, before widespread clinical application is feasible, it will be necessary to test whether cryopreserved ovarian tissue can be transplanted to the oocyte donor, as well as whether it can sustain ovarian steroid-hormone support during her pregnancy⁸.

D. M. Lee, R. R. Yeoman, D. E. Battaglia, R. L. Stouffer, M. B. Zelinski-Wooten, J. W. Fanton, D. P. Wolf

Oregon Health & Science University, Department of Obstetrics and Gynecology, Pediatrics, Physiology and Pharmacology, Oregon National Primate Research Center, Division of Reproductive Sciences, Portland, Oregon 97201, USA
e-mail: leedavi@ohsu.edu

1. Parrott, D. M. *J. Reprod. Fertil.* **1**, 230–241 (1960).
2. Gosden, R. G., Baird, D. T., Wade, J. C. & Webb, R. *Hum. Reprod.* **9**, 597–603 (1994).
3. Schnorr, J. *et al. Hum. Reprod.* **17**, 612–619 (2002).
4. Le Porrier, M., von Theobald, P., Roffe, J. L. & Muller, G. *Cancer* **60**, 2201–2204 (1987).
5. Oktay, K. *et al. J. Am. Med. Assoc.* **286**, 1490–1493 (2001).
6. Duffy, D. M. & Stouffer, R. L. *Hum. Reprod.* **17**, 2825–2831 (2002).
7. Ouhibi, N., Zelinski-Wooten, M. B., Thomson, J. A. & Wolf, D. P. in *Contemporary Endocrinology: Assisted Fertilization and Nuclear Transfer in Mammals* (eds Wolf, D. P. & Zelinski-Wooten, M. B.) 253–284 (Humana, Totowa, New Jersey, 2001).
8. Lutjen, P. *et al. Nature* **307**, 174–175 (1984).

Competing financial interests: declared none.

Supplementary information accompanies the communication on Nature's website.

Tissue engineering

Creation of long-lasting blood vessels

The construction of stable blood vessels is a fundamental challenge for tissue engineering in regenerative medicine. Although certain genes can be introduced into vascular cells to enhance their survival and proliferation, these manipulations may be oncogenic. We show here that a network of long-lasting blood vessels can be formed in mice by co-implantation of vascular endothelial cells and mesenchymal precursor cells, by-passing the need for risky genetic manipulations. These networks are stable and functional for one year *in vivo*.

To create stable vessels, we first seeded human umbilical-vein endothelial cells (HUVECs) and 10T1/2 mesenchymal precursor cells in a three-dimensional fibronectin–type I collagen gel (for methods, see supplementary information). The 10T1/2 cells differentiate into mural cells through heterotypic interaction with endothelial cells^{1–3}.

To permit continuous observation of the engineered vascular networks *in vivo*, we implanted the three-dimensional constructs into mice bearing transparent ‘windows’⁴. The gene for enhanced green fluorescent protein (EGFP) was introduced in order to track the implanted HUVECs⁵.

Initially, the HUVECs formed long, interconnected tubes with many branches that showed no evidence of perfusion (Fig. 1a). Subsequently, they connected to the mouse's circulatory system and became perfused (Fig. 1b; for movie, see supplementary information). This led to a rapid increase in the number of perfused vessels in the first two weeks, followed by stabilization (Fig. 1c), whereas the number of non-perfused vessels gradually decreased and eventually disappeared. By contrast,

constructs prepared from HUVECs alone showed minimal perfusion and had generally disappeared after 60 days (Fig. 1c), even though early morphological changes were similar in the two types of construct.

We labelled 10T1/2 cells with EGFP for *in vivo* microscopy to confirm that they had been incorporated into the vessel wall (Fig. 1d). Cells derived from 10T1/2 cells invested along the perfused vessels. Immunohistochemistry revealed that human cells positive for CD31 (an endothelial cell marker) lined the lumen of the engineered vessels and that these vessels were fortified by 10T1/2-derived cells that expressed the mural-cell marker smooth-muscle α -actin (results not shown).

The 10T1/2 cells that were not associated with new blood vessels did not express appreciable amounts of mural-cell marker. Some mural cells of the engineered vessels derived from the host. However, the engineered vessels in the HUVEC-alone construct rarely survived. Delays in mural-cell recruitment from the underlying host tissue resulted in the regression of most of the vessels engineered from HUVECs alone. In the co-implantation constructs, however,

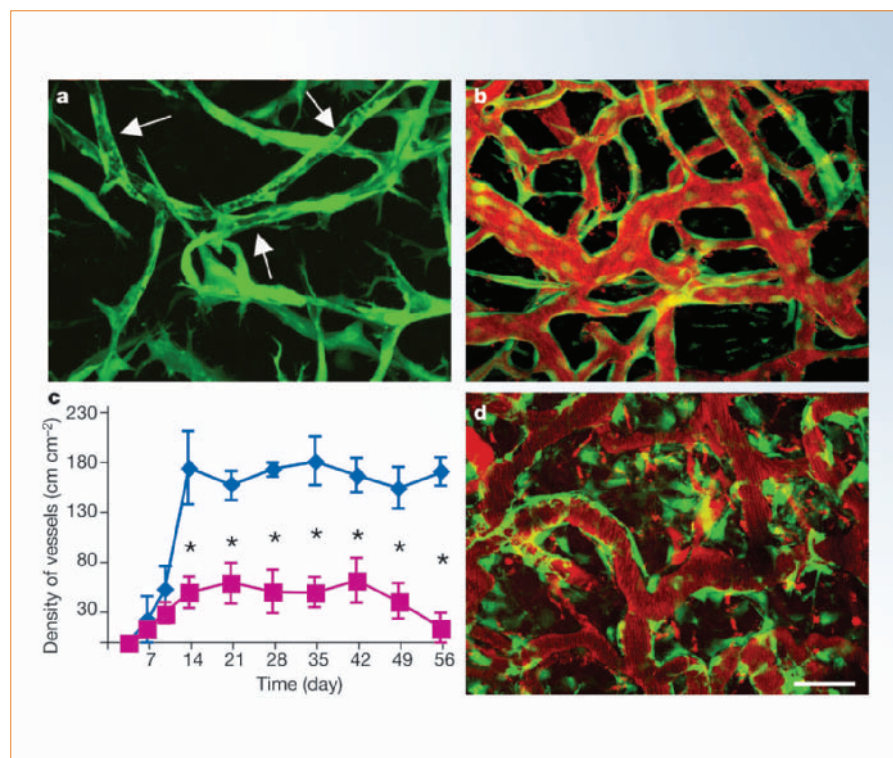


Figure 1 Morphological and functional analysis of engineered blood vessels. Human umbilical-vein endothelial cells (HUVECs) and 10T1/2 mesenchymal precursor cells, or HUVECs alone, were seeded in three-dimensional constructs and then implanted in mice. **a**, **b**, Three-dimensional, intravital, multiphoton laser-scanning microscopy⁴ images of engineered vessels (green, HUVECs expressing enhanced green fluorescent protein (EGFP); red, functional blood vessels contrast-enhanced with Rho-dextran). **a**, Four days after implantation of a HUVEC + 10T1/2 construct. Large vacuoles in the tubes (arrows) resemble the lumen of a capillary vessel but are not perfused (as indicated by the absence of red). **b**, Four months after implantation of a HUVEC + 10T1/2 construct: engineered vessels are still stable and functional. **c**, Temporal changes in functional density of engineered vessels (total length of perfused vessel structure per unit area); $n = 4$; mean $\pm$ s.e.m. is shown; asterisks indicate significance ($P < 0.05$) between the two groups (Mann–Whitney U -test). Top curve, HUVEC + 10T1/2 construct; lower curve, HUVEC-alone construct. **d**, Three-dimensional image of engineered vessels, 4 weeks after implantation of a HUVEC + EGFP-expressing 10T1/2 construct (green, 10T1/2 deprived cells; red, functional blood vessels). Scale bar, 50 μ m. Further details are available from the authors.

the abundance of mural-precursor cells allowed efficient recruitment and investment of mural cells to the engineered vessels, thereby enhancing their stability.

The vascular permeability of the engineered vessels is higher than that of normal, quiescent vessels, but is in the lower range of permeability values induced by various angiogenic molecules (see supplementary information). As in normal microcirculation, arteriolar and venular sides of the engineered vessels are readily identified by the pattern of blood flow. Local administration of a vasoconstrictor, endothelin-1, prompted constriction of the engineered arterial vessels. The few surviving arterioles in the HUVEC-alone constructs were significantly less contractile than those in the combined constructs (see supplementary information). These results indicate that the vessels that formed in the co-implantation construct have a better functionality.

Engineered blood vessels have often been found to be immature and unstable⁶. Genes that enhance the survival and/or proliferation of vascular cells — endothelial cells and mural cells — can be introduced to extend the lifespan of the engineered vessels^{5,7,8}, but these may prove to be oncogenic. We have created long-lasting blood vessels without such genetic manipulation.

In addition to realizing a crucial step in tissue engineering, our system provides a platform for testing the *in vivo* functions of factors that control angiogenesis, vasculogenesis and vessel maturation. The likely existence of endothelial and of smooth-muscle progenitor cells in adults^{9,10} indicates that these cells might serve as a source of autologous cells for engineering blood vessels by using the approach described here.

Naoto Koike, Dai Fukumura, Oliver Gralla, Patrick Au, Jeffrey S. Schechner*, Rakesh K. Jain

Edwin L. Steele Laboratory, Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

e-mail: jain@steele.mgh.harvard.edu

**Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut 06520-8059, USA*

1. Orledge, A. & D'Amore, P. A. *J. Cell Biol.* **105**, 1455–1462 (1987).
2. Hirschi, K. K., Rohovsky, S. A. & D'Amore, P. A. *J. Cell Biol.* **141**, 805–814 (1998).
3. Darland, D. C. & D'Amore, P. A. *Angiogenesis* **4**, 11–20 (2001).
4. Jain, R. K., Munn, L. L. & Fukumura, D. *Nature Rev. Cancer* **2**, 266–276 (2002).
5. Schechner, J. S. *et al. Proc. Natl Acad. Sci. USA* **97**, 9191–9196 (2000).
6. Jain, R. K. *Nature Med.* **9**, 685–693 (2003).
7. Yang, J. *et al. Nature Biotechnol.* **19**, 219–224 (2001).
8. McKee, J. A. *et al. EMBO Rep.* **4**, 633–638 (2003).
9. Yamashita, J. *et al. Nature* **408**, 92–96 (2000).
10. Simper, D., Stalboerger, P. G., Panetta, C. J., Wang, S. & Caplice, N. M. *Circulation* **106**, 1199–1204 (2002).

Competing financial interests: declared none.

Supplementary information accompanies the communication on Nature's website.

COMMUNICATIONS ARISING

Animal behaviour

Inequity aversion in capuchins?

Brosnan and de Waal¹ have shown that capuchin monkeys are more likely to reject a cucumber slice after seeing that another capuchin has received a more attractive grape. In interpreting this finding, the authors make a link to work in humans on 'inequity aversion' and suggest that capuchins, like humans, may reject rewards because they are averse to unequal pay-offs. Here I argue that this interpretation suffers from three problems: the results contradict the predictions of the inequity-aversion model that Brosnan and de Waal cite²; experimental results indicate that humans do not behave like capuchins in similar circumstances; and the available evidence does not suggest that inequity aversion is cross-culturally universal^{3–5}.

I consider these points in turn. Brosnan and de Waal link their findings to work showing that a wide range of experimental behaviour in humans can be understood by introducing a preference for equity into the standard self-interested utility function. The effect of introducing this non-selfish preference is to cause individuals — under certain circumstances — to give up some of their pay-off in order to decrease the gains of other individuals.

Applying the Fehr–Schmidt inequity-aversion model² cited by Brosnan and de Waal to the capuchin experimental situation predicts that capuchins should always eat the cucumber. It does not predict that inequity-averse individuals will reject the food reward, which is what the monkeys did. Rejecting the cucumber increases, not decreases, inequality. Moreover, the grape-receiving capuchins sometimes reached through the cage and stole their partner's discarded cucumber, exacerbating the inequality.

Consistent with inequity aversion in humans, the results from experimental variations of the ultimatum game suggest that humans would not reject a reward unless that rejection reduced the take of the individual who received more. In the ultimatum game, two players are allotted a sum of money to divide. The first player — the 'proposer' — must offer a portion of the sum to the second player — the 'responder' — who must then decide whether to 'accept' or 'reject' the offer. If the responder accepts, he gets the amount of the offer and the proposer receives the remainder. If he rejects, both players get zero. The game is played once, and players never learn their partner's identity. Inequity aversion can explain the willingness of responders to reject low offers (in contrast, pure self-interest predicts that

responders will never reject any non-zero offer). Once some responders are willing to reject low offers, self-interest guarantees that proposers will raise their offers.

Two kinds of variation from the standard ultimatum game suggest that humans, unlike capuchins, would not reject in Brosnan and de Waal's experimental context. Two versions of a reduced-form ultimatum game have been compared⁶: the first was a standard game, except that proposers had only two choices, an equitable allocation or a highly inequitable one; and the second (the 'impunity game') was identical, except that if the responder rejected his offer, the proposer's pay-off remained unchanged (but the responder received zero). In the first version, players were willing to reject inequitable allocations, whereas in the second version (which parallels the capuchin situation), subjects never rejected.

Similar evidence comes from multi-lateral ultimatum games in which one proposer faces multiple responders (see U. Fischbacher, C. M. Fong & E. Fehr, www.iew.unizh.ch/wp/iewwp133.pdf). In this set-up, as long as one of the responders accepts the offer, the proposer gets his pay-off. As before, the responder's ability to affect the proposer's pay-off by rejecting is mitigated by the other responders who might accept. As predicted by inequity aversion, responders decrease their willingness to reject, and proposers drop their offers accordingly. Therefore, although both of these experimental patterns are consistent with inequity aversion, both also seem to be at odds with Brosnan and de Waal's capuchin findings: that is, they show that humans will not reject unless this affects the other's pay-off.

Brosnan and de Waal also suggest that inequity aversion is probably a human universal, and they cite work that uses the ultimatum game in 15 small-scale societies^{3–5}. If responder behaviour (willingness to reject low offers) is taken as the most direct measure of inequity aversion, then our results do not support the universal claim. Although five societies do show evidence consistent with inequity aversion, three others show evidence of trivially little or no inequity aversion. The remainder have so few low offers that no substantial claims can be made.

Joseph Henrich

Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA

e-mail: jhenric@emory.edu

1. Brosnan, S. & de Waal, F. B. M. *Nature* **425**, 297–299 (2003).
2. Fehr, E. & Schmidt, K. M. Q. *J. Econ.* **114**, 817–868 (1999).
3. Henrich, J. *et al.* (eds) *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from 15 Small-Scale Societies* (Oxford University Press, New York, in the press).
4. Henrich, J. *Am. Econ. Rev.* **90**, 973–980 (2000).
5. Henrich, J. *et al. Am. Econ. Rev.* **91**, 73–78 (2001).
6. Bolton, G. E. & Zwick, R. *Games Econ. Behav.* **10**, 95–121 (1995).

Animal behaviour

Fair refusal by capuchin monkeys

Brosnan and de Waal¹ report that capuchin monkeys show evidence of a sense of fairness or 'inequity aversion' because they rejected a less preferred reward when they saw a partner monkey receive a preferred reward for the same task. However, this does not show that monkeys are averse to inequity, only that they reject a lesser reward when better rewards are available. There are risks inherent in seeking anthropomorphic explanations for non-human behaviour.

In the 'inequality test', the monkeys refused to exchange a token for a cucumber slice (non-preferred reward) on 43% of trials when they saw a partner monkey receive a preferred grape reward for the same effort. However, in the 'food control' condition, in which the partner was not present, these same monkeys were just as likely to refuse the cucumber slice when they saw a grape placed where the partner normally sat (49% refusals). There can be nothing inequitable about receiving a non-preferred reward if nobody is receiving anything better. In the food-control condition, the monkeys are refusing the non-preferred reward simply because they can see that a better reward is potentially available. This is therefore the most parsimonious explanation for their refusal to accept the non-preferred reward when they see another monkey receive a better one.

Brosnan and de Waal¹ reject this reward-availability explanation for two reasons. First, in a third condition (the 'effort control' condition), where monkeys saw their partner receive a grape without having to exchange a token, the monkeys were more likely to refuse the cucumber slice than in the food-control condition. On its own, the comparison of the effort-control and food-control conditions is in the direction required by a fairness account. But fairness cannot account for the equally large difference between the effort-control and inequality-test conditions.

The basis of Brosnan and de Waal's second reason for rejecting the reward-availability explanation is in their Fig. 2, which seems to show an increasing trend of non-exchange for the two conditions in which another monkey was present (inequality test and effort control) and a decreasing trend of rejections in the food-control condition where no other monkey was present. Their Fig. 2 shows mean rejections for the first 10 and last 15 trials (not, as stated in the paper, the first 15 and last 10 trials; Brosnan and de Waal, personal communication) averaged across two sessions.

When the cumulative rate of rejections is

represented across all trials of both sessions for Brosnan and de Waal's monkeys, we find that there is no overall increase in rejection rate in the inequality-test and effort-control conditions, and that the rate does not decline across sessions in the food-control condition (results not shown).

Although explanations of animal behaviour in anthropomorphic terms are notoriously prone to imprecision², if 'fairness' or 'inequity aversion' mean anything in this context, they surely imply that individuals reject rewards more often when they see another receive a better reward than when the better reward is simply in view with no one else there to consume it. The very similar levels and patterns of cucumber rejection in the inequality-test and food-control conditions therefore contradict an account based on fairness or inequality.

Clive D. L. Wynne

Department of Psychology, University of Florida,
Box 112250, Gainesville, Florida 32611, USA

e-mail: wynne@ufl.edu

1. Brosnan, S. F. & de Waal, F. B. M. *Nature* **425**, 297–299 (2003).

2. Wynne, C. D. L. *Do Animals Think?* (Princeton Univ. Press, in the press).

Brosnan and de Waal reply — We have shown¹ that animals compare their own rewards with those of others, and accept or reject rewards according to their relative value. Our aim was not to demonstrate that capuchin monkeys make a human response to inequality, but rather to elucidate evolutionary precursors to inequity aversion. We use this term as in ref. 2 — "people resist inequitable outcomes; that is, they are willing to give up some material pay-off to move in the direction of more equitable outcomes" — and specifically focus on "disadvantageous inequity aversion"². The monkeys in our experiment could not change the reward division, and hence could not actively avoid inequality, but we wanted to determine whether they would at least recognize inequality if subjected to it. We found that the capuchins reacted negatively, refusing to complete the interaction.

It is unlikely that inequity aversion appeared *de novo* in humans. It almost certainly evolved because individuals who responded to inequality disadvantageous to themselves increased their relative fitness compared with those who did not. We recognize several potential evolutionary precursors to disadvantageous inequity aversion (S. F. B., H. C. Schiff and F. B. M. de W., manuscript in preparation). First is the ability to recognize that rewards and efforts differ between individuals, which is also required for social learning, a skill present in capuchins³. Second is the propensity to react if another individual receives a better reward for a specific task. Third is sacrifice to alter another individual's outcome.

Our study mainly concerned the second ability, showing that capuchin monkeys react negatively when another individual gets a better reward for the same or less effort on a specific task. This finding suggests that precursors to inequity aversion are present in animals from which our lineage split millions of years ago. Although capuchins may be reacting somewhat differently from adult humans, we have still learned something about the behaviour's possible evolutionary trajectory.

Regarding the cross-cultural study, the lowest mean offer by a proposer in the ultimatum game was 26% of the total, whereas the lowest modal offer was 15%, both by the Machiguenga of Peru⁴. Such relatively high offers would not seem to be consistent with completely selfish individuals who lack any conception of fairness⁵.

As stated earlier¹, although the mere presence of a higher-value reward affects the capuchins, their reaction is not the same as when a conspecific receives the higher-value reward. To ignore the differences between the inequality test and the food-control test is unwarranted. Our Fig. 1 does not permit any conclusions about the effect of the food-control test and was not used for this purpose; it is the data in our Fig. 2 that inspired our claim.

The frequency of refusals across trials increases when a partner receives the reward and decreases when a reward is merely visible. The conservative statistic we chose did not allow significance ($P < 0.05$)¹, but we have since subjected these data to a comparison of the slopes of the linear regressions across trials for each test⁶. This re-analysis shows that refusals in the food-control test decrease across time, whereas those in the inequality test and effort-control condition increase ($F_{2,69} = 28.71$, $P < 0.001$). Our subjects therefore discriminate between a situation in which higher-value food is being consumed by a conspecific and one in which such food is merely visible, intensifying their rejections under only the former condition.

Sarah F. Brosnan, Frans B. M. de Waal

Yerkes National Primate Research Center,
Emory University, Atlanta, Georgia 30329, USA

1. Brosnan, S. F. & de Waal, F. B. M. *Nature* **425**, 297–299 (2003).

2. Fehr, E. & Schmidt, K. M. Q. J. Econ. **114**, 817–868 (1999).

3. Brosnan, S. F. & de Waal, F. B. M. *J. Comp. Psychol.* (in the press).

4. Henrich, J. et al. *Am. Econ. Rev.* **91**, 73–78 (2001).

5. Sigmund, K., Fehr, E. & Nowak, M. A. *Sci. Am.* **83**–87 (2002).

6. Zar, J. H. *Biostatistical Analysis* 3rd edn (Prentice Hall, Saddle River, NJ, 1996).

Corrigendum

When the American sea sturgeon swam east

Arne Ludwig, Lutz Debus, Dietmar Lieckfeldt, Isaac Wirgin, Norbert Bennecke, Ingo Jenneckens, Patrick Williot, John R. Waldman, Christian Pitra
Nature **419**, 447–448 (2002).

The institutional address of A. L. should have been listed as the Institute of Freshwater Ecology and Inland Fisheries and Institute for Zoo and Wildlife Research.

Has the Higgs boson been discovered?

Peter Renton

Department of Physics, Denys Wilkinson Building, University of Oxford, Oxford OX1 3RH, UK

The standard model of particle physics describes the strong and electroweak interactions of fermions (spin-1/2), gauge bosons (spin-1) and a final vital ingredient—the spin-0 Higgs boson, which gives masses to the other particles. But the Higgs boson has yet to be discovered, and its own mass is not specified by the theory. There is some evidence (although statistically not very significant) for its detection at a mass of about $115 \text{ GeV}/c^2$, from electron–positron interactions at LEP (the Large Electron Positron collider). Indirect methods can also be used to constrain the mass of the Higgs boson, because it affects other observable quantities (for example, the mass of the W boson and some measurable properties of the Z boson). An indirect determination of the Higgs boson mass from the most recent measurements of such quantities yields a value compatible with $115 \text{ GeV}/c^2$, but with some important caveats arising from inconsistencies in the present data.

The standard model (SM) of particle physics has developed as a result of painstaking work, over several decades, by a large number of physicists. It is a quantum field theory, combining quantum mechanics and special relativity. It unifies the familiar force of electromagnetism (for example, electricity) with the weak force (such as the reactions that power the Sun) into a unified electroweak force. It also includes the strong force (such as that which binds protons and neutrons to form atomic nuclei). Despite intense experimental effort, no significant deviations from the SM (see ref. 1 for more details) have been observed, so it remains the cornerstone of particle physics.

Table 1 shows the particle content of the SM. The particles are of two main types. First, we have the spin-1/2 fermions, which in turn can be split into leptons (both charged and their neutral neutrino partners) and quarks. Second, we have the spin-1 gauge bosons, which ‘transmit’ the electroweak and strong forces between the fermions. The electromagnetic force is transmitted by the massless photon, whereas the weak force is transmitted by the very massive W and Z bosons. In the underlying theory of the SM there are three W bosons; W^+ and W^- and a neutral W^0 boson, plus another neutral boson called B^0 . However, we observe physically two neutral bosons, the photon and Z boson. These are quantum mechanical ‘mixtures’ of the W^0 and B^0 states. The parameter that describes this mixing is called the ‘weak’ mixing angle, θ_w , and is an important parameter of the model. Measurements of $\sin^2\theta_w$ from a variety of processes give reasonably consistent values, and this consistency is an important test of the SM.

All the 16 particles listed in Table 1 have been discovered. However, with no other ingredients the SM would require them all to be massless, in clear contradiction with nature. The final particle of the SM, the Higgs boson (H), comes to the rescue. It is very important because it is responsible for the mechanism (the so-called Higgs mechanism^{2,3}) by which all other particles acquire mass.

Here we discuss the evidence for the Higgs boson from direct searches and also from an indirect method involving subtle quantum corrections to electroweak quantities; the magnitude of these corrections depends on the mass of the Higgs boson. The consistency of these methods is discussed.

Direct search for the Higgs boson

The search for the Higgs boson is a high priority in particle physics. The Higgs boson is highly unstable and, once produced, decays very quickly to either a fermion–antifermion pair or a pair of bosons. By energy conservation, the Higgs mass, M_H , must be at least twice that of the particle in the pair to which it decays. The mass of the Higgs boson is not specified in the SM, but the strength of the ‘coupling’ of

the Higgs boson to a given particle is proportional to that particle’s mass, so that the Higgs boson decays preferentially to the most massive particles that are allowed by energy and momentum conservation.

The direct search involves first producing the Higgs boson and then detecting the decay products. The mass range that can be explored depends on the energy of the particle collider used. The most powerful and energetic tool available so far in this search is the electron–positron collider, LEP, at CERN in Geneva. Here a search was recently made for the Higgs boson in the interaction $e^+e^- \rightarrow ZH$, with $H \rightarrow b\bar{b}$ (here e^+ indicates positron; e^- , electron; b, quark; $\bar{b}$, antiquark). In the mass range that could be searched for at LEP the decay $H \rightarrow b\bar{b}$ is favoured, as the b quark is the most massive available particle.

A small excess of events, compatible with a Higgs signal, was observed⁴ above the expected background at a mass $M_H \approx 115 \text{ GeV}/c^2$ (here c is the velocity of light). The statistical probability that the background could have fluctuated upwards to the level observed in the data is estimated to be 9%, so the evidence provides only a hint that it could be the elusive Higgs boson. The potential signal is right at the upper mass value accessible to the LEP

Table 1 The standard model

Fermions (spin-1/2)		Gauge bosons (spin-1)		
Leptons	Quarks	Weak	Electromagnetic	Strong
ν_e, ν_μ, ν_τ	u c t (charge 2/3)	W^+, W^-, Z^0	γ	Eight coloured gluons

e^-, μ^-, τ^- d s b (charge $-1/3$)

The charged leptons (e^- , μ^- and τ^-) have the same charge as the electron, e^- , whereas the neutrinos (ν_e, ν_μ, ν_τ) are neutral. The quarks have non-integer charges, as expressed in units of the magnitude of the electron charge. The fermions all have corresponding antiparticles. The u, d, c and s quarks are relatively light, but the b-quark mass is about $5 \text{ GeV}/c^2$ and the t quark is very massive, with a mass of $175 \text{ GeV}/c^2$. In contrast, the lightest charged fermion, the electron, has a mass of $0.5 \text{ MeV}/c^2$. In the simplest version of the SM all the neutrinos are massless, although there is now some evidence that one or more of the neutrinos are not massless. There are three ‘generations’ of leptons and quarks. The first lepton generation is e^- and ν_e , and u and d for quarks. Each successive generation is more massive, giving a mass ‘hierarchy’. The spin-1 gauge bosons transmit the forces between the fermions. The weak force is transmitted by the massive W ($80.4 \text{ GeV}/c^2$) and Z ($91.2 \text{ GeV}/c^2$) bosons. The massless photon transmits the electromagnetic force. The couplings are assumed to be identical for all lepton families (‘lepton universality’). Each of the quarks comes in three different types, or colours. The weak and electromagnetic forces are the same for all colours. The strong force is transmitted by gluons, which transmit the colour charge, and is described by the theory of quantum chromodynamics (QCD). The final ingredient is the (yet to be discovered) spin-0 Higgs boson, which is electrically neutral and gives mass to the other particles. There are a number of ‘constants’ in the SM, whose values must be specified in order to make detailed calculations. There is one constant needed for the QCD part and 17 for the remaining electroweak part. These comprise the masses of the leptons and quarks, and the parameters specifying transitions between quarks. Nearly all of these parameters are associated with the Higgs boson.

collider. Values of M_H lower than $115 \text{ GeV}/c^2$ can be ruled out because the predicted production rate increases rapidly for smaller M_H values, and would have led to a sizeable signal of Higgs events at lower mass values, but these are not observed. A lower mass limit of $114.4 \text{ GeV}/c^2$, at the 95% confidence level, can be set. Lower mass values are excluded with even higher confidence levels.

Indirect search for the Higgs boson

The Higgs boson also manifests itself indirectly through so-called quantum-loop effects. These modify the basic, or lowest-order, electroweak interaction processes. An example of a lowest-order process is shown in Fig. 1. This shows the production of a W boson from the interaction of d and $\bar{u}$ quarks. This is an example of a Feynman diagram, which depicts the propagation in space and time of fundamental particles at the microscopic quantum scale. After the W 'propagates' for a short time it then decays to either a charged lepton plus antineutrino or to either $d\bar{u}$, or $s\bar{c}$, quark pairs. To set the scale, the W boson 'lives' for about 3×10^{-25} seconds, in its own rest frame, before decaying.

There are also electroweak processes beyond the lowest-order diagrams; see Fig. 2. In the top-left diagram of Fig. 2, a W boson (or γ or Z), instead of propagating as just a W, momentarily dissociates into a fermion–antifermion pair. This loop diagram can be regarded as starting with the decay of a W boson into a fermion and antifermion, and then, after a short time, the fermion and antifermion recombine to re-form the W boson, the inverse of the decay process. The probability that this quantum-loop diagram will happen depends on several factors, including the mass of the fermion and antifermion, and also how strongly they 'couple' to the W (or γ or Z). More complicated diagrams involving two or more loops also occur. But calculations show that they add only relatively small numerical corrections to the results obtained using just the simplest loop diagrams.

A W or Z boson propagating through space-time can also momentarily dissociate into a W or Z boson, plus a Higgs boson, as shown in the two lower diagrams in Fig. 2. The effect of these loop diagrams is to change the properties of the W and Z bosons, such as their masses, in a calculable way. For example, the W boson becomes lighter as the Higgs boson becomes heavier. Thus a precise measurement of the W boson mass can be used to set limits on M_H . Measurements of the couplings of the Z boson to both leptons and quarks also give indirect measurements of M_H . Other particles also give loop corrections to these precision electroweak quantities. The largest effects are from the top (t) quark, which is by far the most massive fermion. Indeed, these effects are much larger than those of

the Higgs boson, with the dependence on the top-quark mass M_t being essentially quadratic, compared to a logarithmic dependence on M_H .

Although the value of the Higgs mass is not specified in the SM, some theoretical problems arise if $M_H \geq 1,000 \text{ GeV}/c^2$, where higher-order loops start to dominate and the theory breaks down. So we have a window in which we expect to find the Higgs with a lower limit of $114.4 \text{ GeV}/c^2$, coming from the direct search, and an upper limit from theory of about $1,000 \text{ GeV}/c^2$.

Before the top quark was discovered, in 1994, its existence was 'predicted' by analysis of the precision electroweak data available at that time⁵. Assuming the validity of the SM, these data predicted a top quark with mass $M_t = 178 \pm 11 \text{ GeV}/c^2$, for $M_H = 300 \text{ GeV}/c^2$. (The meaning of the uncertainty, or error, quoted here on a measurement $x \pm \delta x$, is that if the measurement is repeated many times, then the true value will be contained in the range $x - \delta x$ to $x + \delta x$ in 68% of the cases. This is the one-standard-deviation uncertainty, assuming gaussian statistics. This, of course, assumes that both the measured value and the quoted uncertainty are reliably estimated.) Allowing also for the variation in M_H , from $60 \text{ GeV}/c^2$ (the direct-search lower limit at the time) up to $1,000 \text{ GeV}/c^2$, increased the uncertainty on M_t from ± 11 to $\pm 22 \text{ GeV}/c^2$. The mass first measured, after the discovery of the top quark at the Fermilab Tevatron in the USA by the CDF and D0 collaborations^{6,7} was $M_t = 180 \pm 12 \text{ GeV}/c^2$, which is in remarkable agreement with the indirect mass measurement.

Prediction of the Higgs boson mass

The current measured values of those electroweak quantities which are sensitive to the top-quark and Higgs masses are now much more precise. These measurements come from studies of the Z boson properties at LEP and the Stanford Linear Collider (SLAC) in the USA, measurements of the W boson mass at both the Tevatron and CERN and from neutrino–nucleon interactions, together with some other less precise data. The data are now sufficiently precise to constrain the remaining unknown quantity in the SM, namely the Higgs mass. (For a more detailed account of the variables and methods used, together with a recent review of these data, see, for example, refs 8, 9.) Examples of the precision now achieved are that M_Z and M_W (the mass of the Z and the W boson, respectively) are determined to relative accuracies of 2×10^{-5} and 4×10^{-4} , respectively. The weak mixing angle $\sin^2 \theta_w$ is known to a relative precision of 7×10^{-4} .

Fits are made to the ensemble of all these precision electroweak data. The predicted values of these measurements in the SM depend

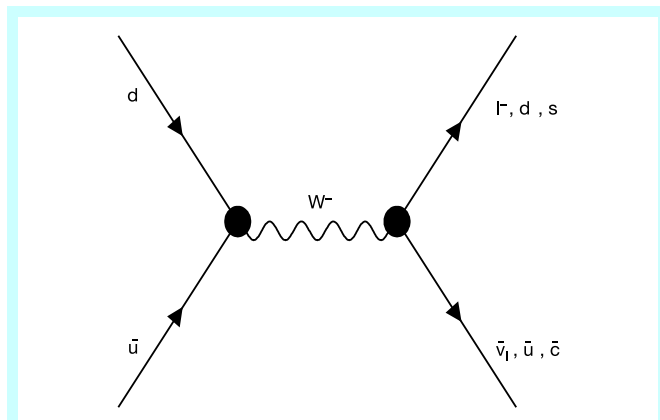


Figure 1 Diagram showing the formation of a W boson from d and $\bar{u}$ quarks. The W propagates in space and time (shown as a wavy line) before decaying into either a charged lepton (l^-) plus antineutrino ($\bar{\nu}_l$) or to quark pairs ($d\bar{u}$ or $s\bar{c}$).

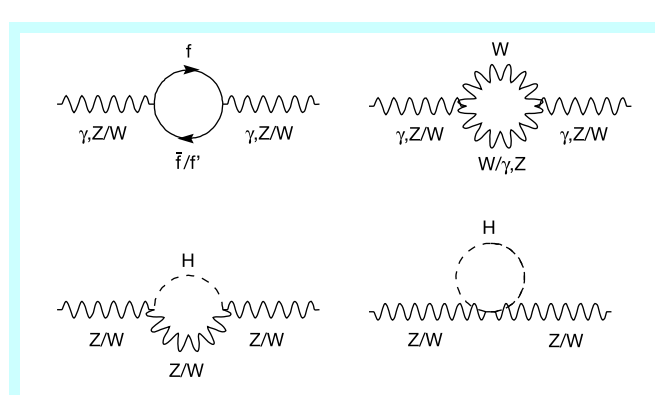


Figure 2 Loop corrections to the propagation of gauge bosons (γ , Z and W). In the top-left diagram, the loop is a fermion–antifermion pair (ff for γ , Z; $\bar{f}f$ for W). In the top-right diagram, the loop is composed of gauge bosons (WW for γ , Z; W γ or Z for W). In the lower diagrams, the loops involve Higgs bosons (ZH for Z; WH for W).

on both M_t and M_H , as well as on the values of some well-known 'constants' of the SM. For each measurement, the measured value, plus the one-standard-deviation uncertainty, is used (for example, $\Gamma_Z^{\text{meas}} \pm \delta\Gamma_Z$ for the total width of the Z boson). A χ^2 function is formed for all these electroweak data (for example, a contribution $(\Gamma_Z^{\text{meas}} - \Gamma_Z^{\text{SM}})^2 / \delta\Gamma_Z^2$ from Γ_Z). In forming the total χ^2 , any correlations between the measurements are taken into account. Measurements of 20 separate quantities are used. Most of these measurements are, in fact, averages of several different determinations of the same physical quantity. The χ^2 function is minimized with respect to M_t and M_H , and the best-fitted values of M_t and M_H are extracted.

Figure 3 shows the results of the indirect determination of M_H from all the data, as well as some other data sets that are discussed later. The direct top-quark mass measurement of $M_t = 174.3 \pm 5.1 \text{ GeV}/c^2$ is used in this fit. This is obtained from the combined results of the CDF and D0 experiments^{10–12}. The fitted value for the Higgs mass is $M_H = 96^{+60}_{-38} \text{ GeV}/c^2$. The quoted errors are asymmetric because the dependence of the predicted values of the measured quantities on M_H is logarithmic (the errors on $\log(M_H)$ are, in fact, reasonably symmetric. However, when transformed into M_H values the errors become asymmetric, with masses above the best-fitted value being less well constrained than those below.). The quoted uncertainties correspond to a change in χ^2 with unity, with respect to the minimum value. The variation of χ^2 with

M_H can, using the properties of χ^2 , be used to set an upper limit on M_H . This gives $M_H \leq 219 \text{ GeV}/c^2$ at the 95% confidence level. The estimated uncertainty on the theoretical predictions of the SM, due to uncalculated higher-order terms, is taken into account.

If the direct top-quark mass measurement is excluded, then the fit gives $M_t = 174^{+11}_{-8} \text{ GeV}/c^2$ and $M_H = 92^{+130}_{-48} \text{ GeV}/c^2$. This shows that this indirect top-quark mass estimate is in excellent agreement with the directly measured value. It also shows the importance of improving the precision of the direct top-quark mass measurement, as it substantially reduces the uncertainty on M_H .

Thus the data appear to support the existence of a Higgs boson with a mass close to the possible signal observed in the direct search at LEP, around $115 \text{ GeV}/c^2$. So the situation is somewhat analogous to that of the prediction and discovery of the top quark in 1994, except that for the Higgs boson the confidence level for the direct observation is rather weak.

The caveats

There are, however, some potential problems with these SM fits. The overall probability of the fit to all the data is only about 5%. This probability is estimated from the minimum χ^2 value and the number of degrees of freedom in the fit. The meaning is that if the ensemble of measurements were repeated over and over many times, then in only 5% of the cases would the agreement with the SM be worse. Thus the data are not all very compatible with the SM, but the level of disagreement is not sufficient to conclude that the SM is incorrect.

The individual measurement with the worst agreement with the SM prediction is that from the NuTeV neutrino–nucleon interaction experiment¹³, which gives a contribution of 8 to the total minimum fit χ^2 , of 25. If this measurement is removed from the fit then the fitted Higgs mass is reduced by only $5 \text{ GeV}/c^2$, but the probability increases from 5% to 28% (Fig. 3). Thus this measurement carries very little weight in the fit, but substantially worsens the fit probability. The different measured quantities used in the SM fits are sensitive to different possible violations of the SM. So the

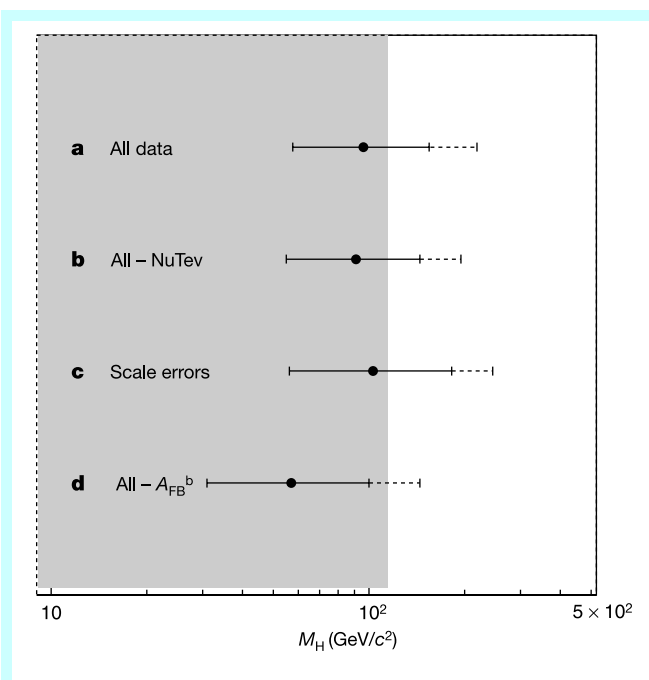


Figure 3 Results of the indirect determination of M_H . For each fit, the large dot shows the best-fitted value and the solid line the one-standard-deviation uncertainties. The dashed line gives the 95% one-sided confidence level upper limit. The data come from LEP, SLAC, the Tevatron and neutrino–nucleon interactions. The shaded region is that excluded at the 95% confidence level by direct-search results. **a**, The fit to all electroweak data, giving $M_H = 96^{+60}_{-38} \text{ GeV}/c^2$ and $M_H \leq 219 \text{ GeV}/c^2$ at the 95% confidence level. The probability for the fit is 5%. **b**, The result if the NuTeV data are excluded from the fit. The probability increases to 28%, and the 95% upper limit becomes $195 \text{ GeV}/c^2$. In fit **c**, the errors of the quantities most sensitive to M_H , and which are not themselves very self-consistent (as can be seen in Fig. 4), are increased by a scale factor. The best-fitted value increases by $5 \text{ GeV}/c^2$ with respect to fit **a**, and the 95% upper limit increases to $246 \text{ GeV}/c^2$. In fit **d**, the forward–backward asymmetry for b quarks A_{FB}^b is removed and the fit probability becomes 17%. The best-fit value is $M_H = 57^{+44}_{-26} \text{ GeV}/c^2$ and the 95% upper limit is $143 \text{ GeV}/c^2$. There is poor agreement with the results of the direct search.

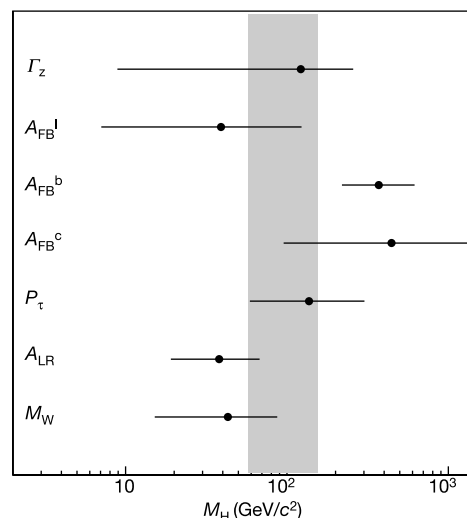


Figure 4 The Higgs mass values obtained from the most sensitive individual electroweak measurements. These quantities are the Z boson total width Γ_Z , the forward–backward asymmetries for leptons (A_{FB}^l) and b and c quarks (A_{FB}^b and A_{FB}^c) measured in Z boson decays with respect to the incident electron direction, the τ -lepton polarization P_τ , the asymmetry in Z decays obtained by using left- and right-handed polarized beams A_{LR} , and the mass of the W boson M_W . In each fit the other parameters of the SM are fixed. The error bars shown correspond to one standard deviation. The vertical band shows the one-standard-deviation limits of the fit to all the data.

NuTeV result may indicate new physics beyond the SM, but this is not statistically compelling at present.

Out of the 20 electroweak measurements used in the fit, most of the sensitivity to M_H comes from a subset of seven. The values of M_H obtained from each of these quantities individually are shown in Fig. 4. They show a wide range and are not very compatible, with A_{FB}^b and A_{FB}^c favouring a more massive value of M_H than the other quantities. (A_{FB}^b and A_{FB}^c are the forward-backward asymmetries for b and c quarks, respectively.)

Measurements of these seven quantities contribute 13 to the total χ^2 of 25. The probability that they are compatible is only about 4%. The Particle Data Group¹⁴, which makes combinations of a variety of measurements in particle physics, argues that historically a large χ^2 is often indicative of systematic uncertainties in some, or all, measurements being underestimated. A possible procedure in this situation is to scale the total quoted uncertainties on these quantities by a factor $\sqrt{\chi^2/(n-1)}$, where n is the number of measurements. Here we apply this scaling factor to just these quantities. If this is done (with a factor 1.5), the fitted Higgs mass becomes $M_H = 103^{+77}_{-47}$ GeV/ c^2 , with a fit probability of 22% for all the electroweak data (Fig. 3). However, it is worth noting that the ratios of the quoted systematic to statistical uncertainties on these quantities are typically around 50–60%, so this would require a very large underestimation of the systematic uncertainties—and this seems unlikely.

The measurement with the second largest contribution to χ^2 in the fit to all the data is A_{FB}^b . If this measurement alone is removed from the fit then the fit probability becomes 17% and the fitted Higgs mass becomes $M_H = 57^{+44}_{-26}$ GeV/ c^2 . This value is in poor agreement with the direct-search lower limit of 114.4 GeV/ c^2 (Fig. 3). The confidence level from this fit that the Higgs mass is larger than this direct-search limit is only 11%, compared to 36% when all the data are fitted.

Where do we stand?

Despite the problems outlined above, the χ^2 fit to the data nevertheless shows a very significant dependence on M_H , and favours a value well below 1,000 GeV/ c^2 (Fig. 3). The poor χ^2 values of the fits may indicate either that there are sizeable statistical fluctuations in the data or that some of the measurements are prone to systematic effects that have not been understood. The current data are reasonably compatible with $M_H \approx 115$ GeV/ c^2 . However, somewhat larger mass values are also compatible. The data suggest that either there is a Higgs boson around this mass or that there are some new phenomena that mimic the effects of the Higgs boson on the electroweak data.

The SM has various problems. It is clearly incomplete, because it does not include the force of gravity. It does not explain why there are three generations of leptons and quarks and why there is a mass ‘hierarchy’, with successive generations being heavier. Although the Higgs boson is a necessary ingredient of the model, it also causes theoretical problems. If M_H is calculated in the SM, then the loop diagrams lead to a quadratically divergent result. This problem could be ‘cured’ if successively more complicated loop diagrams largely cancelled each other out, to leave a finite result for M_H . But such behaviour of the theory does not seem ‘natural’.

An example of new physics beyond the SM, which offers a potential solution is supersymmetry, in which there are several

Higgs bosons together with so-called super-partners for all the SM particles. These super-partners essentially cancel the divergences, leaving a finite value for the Higgs mass. But the interpretation including supersymmetry is complicated, as there are more than 100 extra parameters introduced. Although the introduction of supersymmetry can improve some aspects of the fits, it does not lead to a natural explanation of either the NuTeV result or the possible anomalies in the b-quark asymmetry.

The indirect limits will be improved using data from the ongoing Run 2 at the Tevatron, aimed at reducing the uncertainties on the masses of the W boson and top quark, each by a factor of about two. The Large Hadron Collider (LHC) at CERN, starting in 2007, and any future e^+e^- collider, will also further improve these measurements.

But at present we can only conclude that the data are reasonably consistent with the SM, provided that the Higgs boson is in the mass range of about 100–200 GeV/ c^2 . The inconsistencies in the present data prevent a more precise conclusion. The indirect method for M_H gives sensitivity to an energy region not directly accessible with current facilities. But it is not a substitute for direct experimental observation. It tells us only that there is a Higgs boson, or some other physical phenomena that mimic some of the Higgs properties. A conclusive direct search for the Higgs boson at the Tevatron seems to be beyond the projected capabilities of this machine. However, the LHC will open up this entire energy regime to direct searches, and should resolve the question of whether the Higgs boson exists, or, if not, what is the new physics behind the generation of mass. □

doi:10.1038/nature02324.

1. Renton, P. B. *Electroweak Interactions* (Cambridge Univ. Press, Cambridge, 1990).
2. Higgs, P. W. Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132–133 (1964).
3. Higgs, P. W. Spontaneous symmetry breaking without massless bosons. *Phys. Rev.* **145**, 1156–1163 (1966).
4. ALEPH, DELPHI, L3 and OPAL Collaborations. Search for the Standard Model Higgs Boson at LEP. *Phys. Lett. B* **565**, 61–75 (2003).
5. The LEP Collaborations, ALEPH, DELPHI, L3 and OPAL. *Combined Preliminary Data on Z Parameters from the LEP Experiments and Constraints on the Standard Model* (CERN/PPE/94–187 CERN, Geneva, 1994).
6. Abe, F. *et al.* Observation of top quark production in pp collisions with the collider detector at Fermilab. *Phys. Rev. Lett.* **74**, 2626–2631 (1995).
7. Abachi, S. *et al.* Observations of the top quark. *Phys. Rev. Lett.* **74**, 2632–2637 (1995).
8. Renton, P. B. Precision electroweak tests of the Standard Model. *Rep. Prog. Phys.* **65**, 1271–1330 (2002).
9. Gambino, P. The top priority: Precision electroweak physics from low to high energy. *Proc. Lepton Photon 2003* (in the press).
10. Affolder, T. *et al.* Measurement of the top quark mass with the collider detector at Fermilab. *Phys. Rev. D* **63**, 032003 (2001).
11. Abbott, B. *et al.* Measurement of the top quark mass in the dilepton channel. *Phys. Rev. D* **60**, 052001 (1999).
12. Demortier, L. *et al.* Combining the top quark mass results for Run 1 from CDF and D0. (FERMILAB-TM-2084, Fermilab, Batavia, IL, 1999); (<http://tevewwg.fnal.gov/top/fermilab-tm-2084/pdf>) (1999).
13. Zeller, G. P. *et al.* Precision determination of electroweak parameters in Neutrino-Nucleon Scattering. *Phys. Rev. Lett.* **88**, 091802 (2002).
14. Hagiwara, K., Particle Data Group *et al.*. *Phys. Rev. D* **66**, 010001 (2003).

Acknowledgements I thank my colleagues in the LEP Electroweak Working Group for discussions.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (p.renton1@physics.ox.ac.uk).

Germline stem cells and follicular renewal in the postnatal mammalian ovary

Joshua Johnson*, Jacqueline Canning*, Tomoko Kaneko, James K. Pru & Jonathan L. Tilly

Vincent Center for Reproductive Biology, Vincent Obstetrics and Gynecology Service, Massachusetts General Hospital, and Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, Massachusetts 02114, USA

* These authors contributed equally to this work

A basic doctrine of reproductive biology is that most mammalian females lose the capacity for germ-cell renewal during fetal life, such that a fixed reserve of germ cells (oocytes) enclosed within follicles is endowed at birth. Here we show that juvenile and adult mouse ovaries possess mitotically active germ cells that, based on rates of oocyte degeneration (atresia) and clearance, are needed to continuously replenish the follicle pool. Consistent with this, treatment of prepubertal female mice with the mitotic germ-cell toxicant busulphan eliminates the primordial follicle reserve by early adulthood without inducing atresia. Furthermore, we demonstrate cells expressing the meiotic entry marker synaptonemal complex protein 3 in juvenile and adult mouse ovaries. Wild-type ovaries grafted into transgenic female mice with ubiquitous expression of green fluorescent protein (GFP) become infiltrated with GFP-positive germ cells that form follicles. Collectively, these data establish the existence of proliferative germ cells that sustain oocyte and follicle production in the postnatal mammalian ovary.

Although males retain germline stem cells (GSCs) for spermatogenesis throughout adult life, oocyte production in females of most mammalian species is believed to cease before birth^{1–5}. Accordingly, a central dogma of mammalian reproductive biology is that females are born with a finite, non-renewing pool of germ cells, all of which are arrested in meiosis I (oocytes) and are enclosed by somatic cells in structures referred to as follicles^{1–5}. Oocyte numbers decline throughout postnatal life^{6–8} through mechanisms involving apoptosis^{9,10}, eventually leaving the ovaries barren of germ cells¹¹. In humans, exhaustion of the oocyte reserve occurs around the fifth decade of life, driving the menopause¹². The process that is believed to occur in female mammals with respect to germ-cell development differs from that of several invertebrate organisms, including *Drosophila melanogaster*, in which GSCs maintain oocyte production in adult ovaries^{13–15}. This apparent evolutionary disparity is female-specific, as the role of GSCs in maintaining spermatogenesis has been documented in essentially all metazoan species^{13–15}.

Discordance in follicle loss versus atresia

To better define germ-cell dynamics in female mammals, a study was initially undertaken to assess the numbers of healthy (non-atretic) and degenerating (atretic) follicles in ovaries of C57BL/6 mice. Analysis of non-atretic quiescent (primordial) and early growing (primary, preantral) follicle numbers revealed that approximately one-third of the peak endowment of immature follicles was lost during development to young adulthood (Fig. 1a), consistent with past studies of follicle depletion in mice⁷. However, unlike all previous studies of follicle development in mammals that we are aware of, a parallel assessment of immature follicle atresia was conducted and compared with the numbers of non-atretic follicles lost over time. Through the first 20 days of age, atresia occurred at a low but constant rate (Fig. 1b), consistent with a proportional decline in non-atretic follicle numbers during this time period (Fig. 1a). However, the incidence of atresia increased markedly by day 30 and further by day 40, reaching a peak level of more than 1,200 dying follicles per ovary on day 42 that was maintained well into reproductive life (Fig. 1b).

Clearance of apoptotic cells *in vivo* occurs within 3–18 h^{16–18}. Nonetheless, experiments were conducted to rule out the possibility that the large atretic follicle population observed in adult animals simply represented accumulation of oocyte corpses in follicles that had degenerated weeks earlier. The first experiment, based on past studies showing that extensive levels of oocyte apoptosis occur in the newborn mouse ovary coincident with follicle formation^{19,20}, evaluated changes in the number of non-atretic oocytes between days 1 and 4 postpartum compared with the number of degenerative oocytes on day 4. More than 8,000 non-atretic oocytes were present per ovary on day 1, and this pool was reduced by almost 50% by day 4 (Supplementary Table S1). However, only 218 degenerative oocytes per ovary were found on day 4, indicating that over 3,300 oocytes had died and been cleared from the ovary between days 1 and 4 postpartum (Supplementary Table S1). The second approach to assess clearance rates of degenerative oocytes used the chemical 9,10-dimethylbenz[a]anthracene (DMBA) to synchronize primordial and primary follicle atresia. Past studies have shown that DMBA induces degeneration of immature oocytes in a manner that morphologically resembles developmental oocyte death²¹. In addition, the pro-apoptotic Bax protein is required for immature follicle atresia under normal conditions⁹ and in response to DMBA exposure²², further underscoring the similarities between the two processes. In female mice given a single injection of DMBA, the incidence of follicle atresia increased markedly between 24 and 48 h after injection, and remained at a plateau of approximately 850 atretic follicles per ovary from 48 to 72 h after injection (Fig. 1c). By 96 h after injection, there were no healthy primordial or primary follicles remaining in the ovaries, and the incidence of atresia returned to near-basal levels (Fig. 1c). Therefore, similar to the clearance rate of degenerative oocytes between postnatal days 1 and 4 discussed earlier, DMBA-induced synchronization of atresia revealed that over 3,500 oocytes contained within primordial and primary follicles initiated apoptosis and were cleared from the ovary within a 3-day period.

Given this, the finding that from 1% (days 8, 12 and 20) to as much as 16% (day 40) or more (33%, day 42; Fig. 1d) of the

immature follicle pool is degenerating at any given time under normal conditions would be predictive of complete exhaustion of the follicle reserve by young adulthood. However, the non-atretic pool of follicles declined from peak levels on day 12 by only 36% on day 40 (Fig. 1a). This indicated that the rate of follicle depletion during postnatal life, as determined by assessing changes in non-atretic follicle numbers, was highly incongruous with the numbers of follicles actually being eliminated from the ovaries through atresia in the same time frame. To confirm that these findings were not a phenomenon related to C57BL/6 mice, changes in follicle numbers from birth to adulthood were analysed in other strains of mice, and compared with corresponding data from C57BL/6 females. In CD1 mice, the non-atretic follicle pool declined by only 4% between days 4 and 42 postpartum, despite a relatively high incidence of atresia, comparable to that observed in C57BL/6 females (Fig. 1d). Even more striking, the non-atretic follicle population in AKR/J mice was 20% larger on day 42 than on day 4, again despite a marked incidence of atresia (Fig. 1d). These data, which extend past work showing that genetic strain has a marked impact on female germ-cell dynamics²³, highlight a clear discordance between changes in non-atretic follicle numbers and the corresponding incidence of atresia in the postnatal mammalian ovary.

Postnatal female germ-cell proliferation

In the light of these data, the possibility that oocyte and follicle renewal are still ongoing in the postnatal mouse ovary was examined by several approaches. First, histological analysis of juvenile and young adult ovaries revealed the presence of large ovoid cells, resembling germ cells of fetal mouse ovaries^{24,25}, in the surface epithelial cell layer covering the ovary (Fig. 2a). However, these cells were not intimately associated with smaller squamous epithelial cells in any type of structure reminiscent of a follicle. Immunohistochemical staining for mouse Vasa homologue (MVH), a gene expressed exclusively in germ cells of both vertebrate and invertebrate species^{26,27}, confirmed that these large ovoid cells were of a

germline lineage (Fig. 2b). To document the proliferative potential of these MVH-positive cells, juvenile and young adult female mice were injected with 5-bromodeoxyuridine (BrdU) and ovaries were collected 1 h later for dual immunostaining analysis of BrdU incorporation and MVH expression. These experiments confirmed the presence of BrdU–MVH double-positive cells, all of which were found in the ovarian surface epithelium (Fig. 2d–f). This finding was not the result of BrdU incorporation associated with mitochondrial DNA replication or DNA repair in post-meiotic female germ cells, as the degree of BrdU incorporation observed in cells due to either of these processes is several log orders less than that seen during replication of the nuclear genome during mitosis^{28,29}. Consistent with this, the BrdU immunoreaction in MVH-positive cells was comparable in intensity to that observed in replicating follicular somatic (granulosa) cells, and, more importantly, was never observed in adjacent oocytes contained within follicles (Fig. 2d, e). Analysis of ovaries immunostained for MVH and counterstained with propidium iodide confirmed the presence of germ cells in various stages of mitosis (Fig. 2g, h). These data, taken with the histomorphometric findings discussed earlier, build a strong case for germ-cell proliferation and follicle renewal in the postnatal mouse ovary.

Meiotic entry in postnatal ovaries

Replication of germ cells to produce oocytes for follicle formation in postnatal life would require expression of genes involved in the initiation of meiosis. Thus, expression of synaptonemal complex protein 3 (SCP3), a meiosis-specific protein needed for formation of axial lateral elements of the synaptonemal complex^{30,31}, was examined in juvenile and young adult mouse ovaries. Immunohistochemical localization of SCP3 revealed individual immunoreactive cells in or proximal to the surface of the ovary (Fig. 3a, b). The possibility that SCP3 was simply carried over as a stable protein product in oocytes formed during the perinatal period was ruled out by the finding that oocytes contained within immature follicles were not immunoreactive (Fig. 3c, d). Postnatal ovarian expression

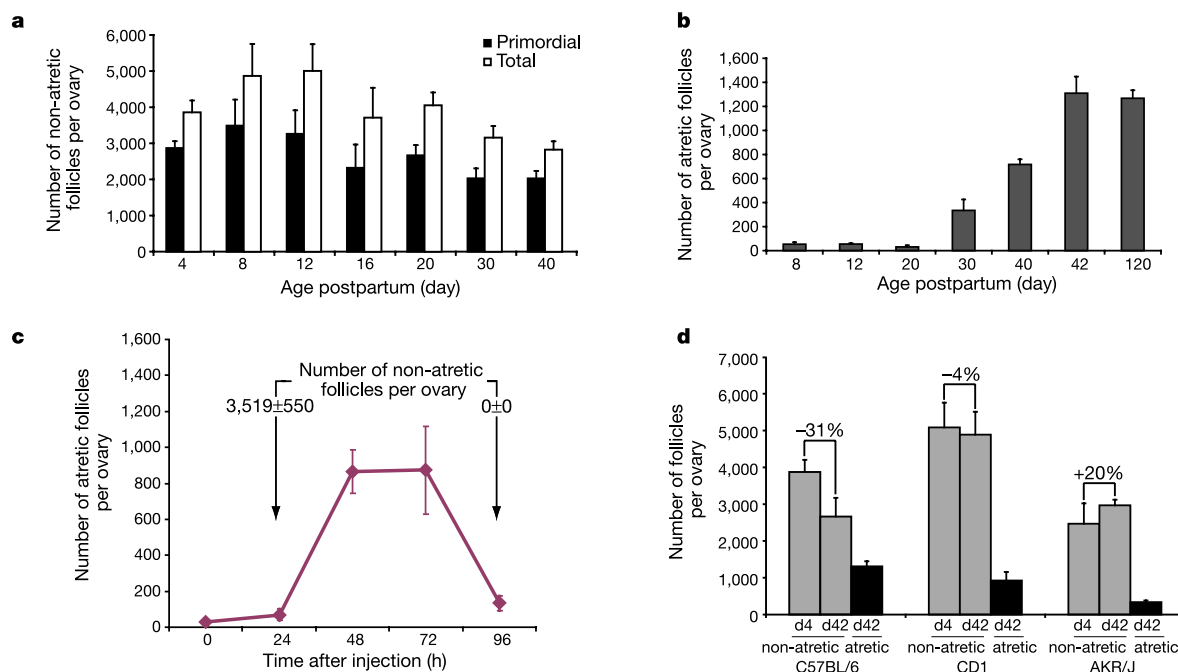


Figure 1 Postnatal ovarian germ-cell dynamics. **a**, **b**, Numbers of non-atretic (**a**) and atretic (**b**) primordial and total immature (primordial, primary, small preantral) follicles in mouse ovaries during postnatal development (**b**, total immature follicles; mean $\pm$ standard error, $n = 3$ –4 mice per age group). **c**, Incidence of primordial and

primary follicle atresia in ovaries exposed to DMBA on day 25 postpartum (mean $\pm$ standard error, $n = 4$ –5 mice per time point). **d**, Comparison of non-atretic and atretic immature follicle numbers in C57BL/6, CD1 and AKR/J strains of mice (mean $\pm$ standard error, $n = 3$ –5 mice per age group).

of *Scp3* was confirmed at the messenger RNA level, as was expression of the endonuclease *Spo11* and the recombinase *Dmc1* (Fig. 3e), both of which are also required for the initiation of meiosis in mammals³². Past work has shown that expression of *Scp3* and *Dmc1* in germ cells is restricted to the zygotene or pachytene stages of meiosis^{30–32}. As these stages are earlier than the late diplotene stage where the first meiotic arrest in oocytes is observed³³, the presence of these mRNA transcripts, like *SCP3* protein, cannot be considered simply as a reflection of oocytes being in meiosis. The levels of *Scp3*, *Spo11* and *Dmc1* expression in adult ovaries ranged from 6% (*Spo11*, *Dmc1*) to 25% (*Scp3*) of those observed in adult testes (Fig. 3f), which is significant considering that daily postnatal germ-cell output in the testis far exceeds that estimated for the ovaries (see next section). Ovarian expression of all three meiosis-related genes declined with age (Fig. 3e), and minimal to no expression of these genes was observed in non-gonadal tissues (Fig. 3g).

Primordial follicle renewal

The importance of proliferative germ cells to replenishment of the postnatal follicle pool was further verified by the use of busulphan, a germ-cell toxicant widely used in spermatogonial

stem-cell characterization in male mice^{34–36}. In the testis, busulphan specifically targets GSCs and spermatogonia, but not post-meiotic germ cells, leading to spermatogenic failure³⁴. Female rodents exposed *in utero* show a similar gametogenic failure in response to busulphan only if the chemical is given during the window of fetal ovarian germ-cell proliferation, as females exposed to busulphan *in utero* after germ-cell proliferation has ceased are born with ovaries that are histologically and functionally similar to ovaries of vehicle-exposed mice³⁷. In the present experiments, female mice were injected with vehicle or busulphan on day 25 and again on day 35, and ovaries were collected 10 days after the second injection to analyse changes in non-atretic primordial follicle numbers. Ovaries of females treated with busulphan possessed less than 5% of the primordial follicle pool present in vehicle-treated controls 20 days after the start of the experiment (Fig. 4a). However, busulphan-exposed ovaries retained an otherwise normal histological appearance, including the presence of healthy maturing follicles with non-degenerative oocytes, as well as corpora lutea, indicative of ovulation (Fig. 4b–e).

A previous investigation of adult female mice exposed to busulphan reported a similar decline in immature follicle numbers with complete oocyte loss occurring sometime between 2–17 weeks after injection³⁸. The depletion of follicles was attributed to cytotoxic actions of busulphan in follicle-enclosed oocytes, although the incidence of atresia was not directly evaluated³⁸. If busulphan causes ovarian failure in adult mice because of toxicity to existing oocytes, we found it puzzling that the time frame for oocyte loss after this insult would be weeks to months rather than hours to days, as is the case with other germ-cell toxicants such as DMBA (Fig. 1c) or cyclophosphamide³⁹. To clarify whether loss of primordial follicles observed in busulphan-treated females (Fig. 4a) results from toxicity to existing oocytes, ovaries were collected from female mice at multiple points during and after the busulphan dosing regimen described above, and they were analysed for the incidence of primordial follicle atresia. Busulphan caused a slight, transient increase in the number of atretic primordial follicles, with a plateau of only 46 per ovary 5 days after the first injection that quickly declined to basal levels thereafter (Fig. 4a, inset). This relatively minor and acute atretic response to busulphan was negligible considering that over 2,000 primordial follicles were absent in busulphan-exposed ovaries compared with vehicle-treated controls

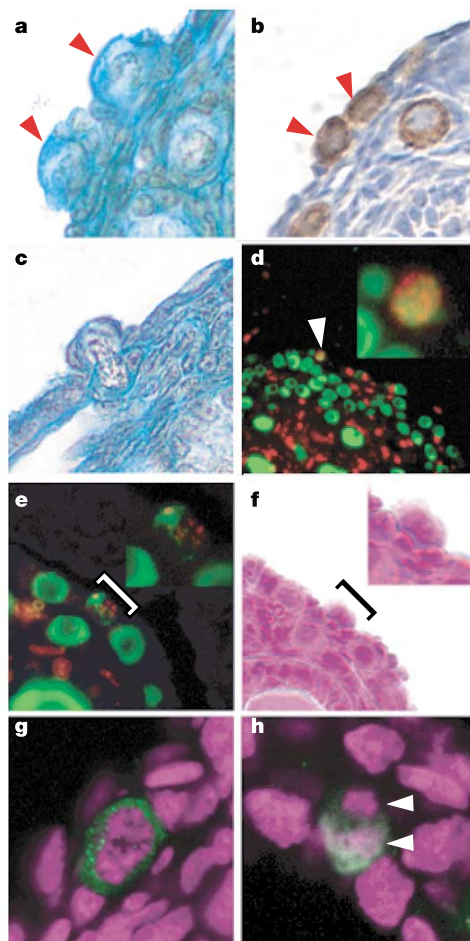


Figure 2 Germ-cell proliferation in juvenile and young adult ovaries. **a**, Presumptive GSCs (arrowheads) in the surface epithelium. **b**, Immunohistochemical detection of MVH (brown reaction product) in presumptive GSCs (arrowheads). **c**, Presumptive GSC undergoing mitosis across the surface epithelium. **d**, **e**, Dual immunostaining for BrdU (red) and MVH (green) in juvenile and young adult mouse ovaries. **f**, Haematoxylin and eosin staining of the section shown in **e**, highlighting the dividing germ cell. **g**, **h**, Chromatin morphology (propidium iodide counterstaining, violet) within MVH-positive cells (green) shows a germ cell in prometaphase (**g**) and a germ cell in metaphase (**h**; sister chromatids highlighted by arrowheads).

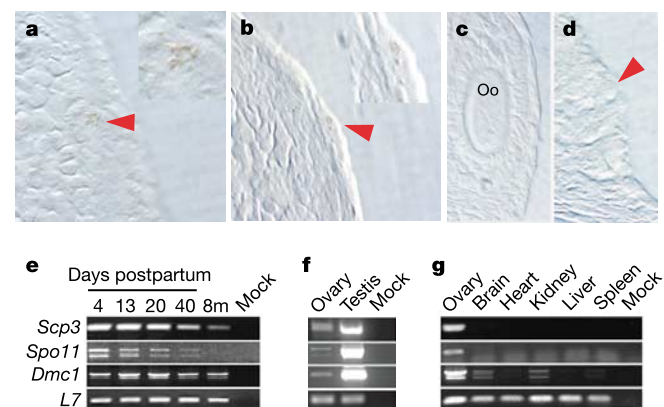


Figure 3 Meiotic entry gene expression in postnatal ovaries. **a–d**, SCP3 immunostaining (brown) in single cells (**a**, **b**) located in or proximal to the surface epithelium of juvenile and young adult mouse ovaries (**c**, small preantral follicle; **d**, primordial follicle with oocyte marked by arrowhead). Oo, oocyte. **e**, Expression of *Scp3*, *Spo11* and *Dmc1* in ovaries collected at the indicated days of age or at 8 months (8m) postpartum. *L7*, ribosomal gene co-amplified as an internal control; mock, mock-transcribed ovarian RNA samples. **f**, **g**, *Scp3*, *Spo11* and *Dmc1* expression in ovaries versus testes (**f**) or in various tissues (**g**) collected from young adult mice.

(Fig. 4a). These data reinforce the concept that proliferative germ cells not only persist in the postnatal ovary but also are required to routinely renew the follicle pool.

To determine the rate of primordial follicle renewal in the postnatal mouse ovary, we evaluated our findings in the context of a past investigation of the kinetics of follicle maturation in female mice⁷. Faddy and colleagues⁷ demonstrated that the primordial follicle pool is decreased on average by 89 follicles per day, owing to either degeneration or growth activation to the primary stage of development, between days 14 and 42 postpartum. In our study's comparable window of time (day 16–40 postpartum; Fig. 1a), this rate of exit would be expected to reduce the primordial follicle population by 2,136 follicles over this 24-day period. However, the number of primordial follicles declined by only 294 between days 16 and 40 postpartum (Fig. 1a). The difference between these two values—or 1,842 primordial follicles—represents the rate of

primordial follicle renewal over this 24-day period, yielding an average of 77 new primordial follicles per ovary per day. If this calculation is accurate, then the rate of primordial follicle depletion per day should be the difference between the rate of exit per day provided by Faddy and co-workers⁷ (89 follicles) and the rate of renewal per day (77 follicles), for a net loss of 12 primordial follicles per ovary per day. Using this value, the primordial follicle pool would be expected to decline between days 16 and 40 postpartum by a total of 288 follicles, a number very close to that derived from comparing the actual counts of non-atretic primordial follicles on day 16 versus day 40 (2,334 versus 2,040, or a net loss of 294 primordial follicles; Fig. 1a).

Folliculogenesis and germline stem cells

In a final set of experiments, transgenic mice with ubiquitous expression of GFP⁴⁰ were used to provide additional evidence for ongoing folliculogenesis in postnatal life. Ovarian fragments harvested from adult wild-type mice were grafted into the ovarian bursa cavity of transgenic female siblings after removal of approximately one-half of the host's ovary. After 3–4 weeks, ovaries were collected and processed for GFP expression. The grafted ovarian fragments, upon gross visual inspection, showed evidence of neo-vascularization and adhesion to the host ovarian tissue (Supplementary Fig. S1). Confocal microscopic analysis revealed

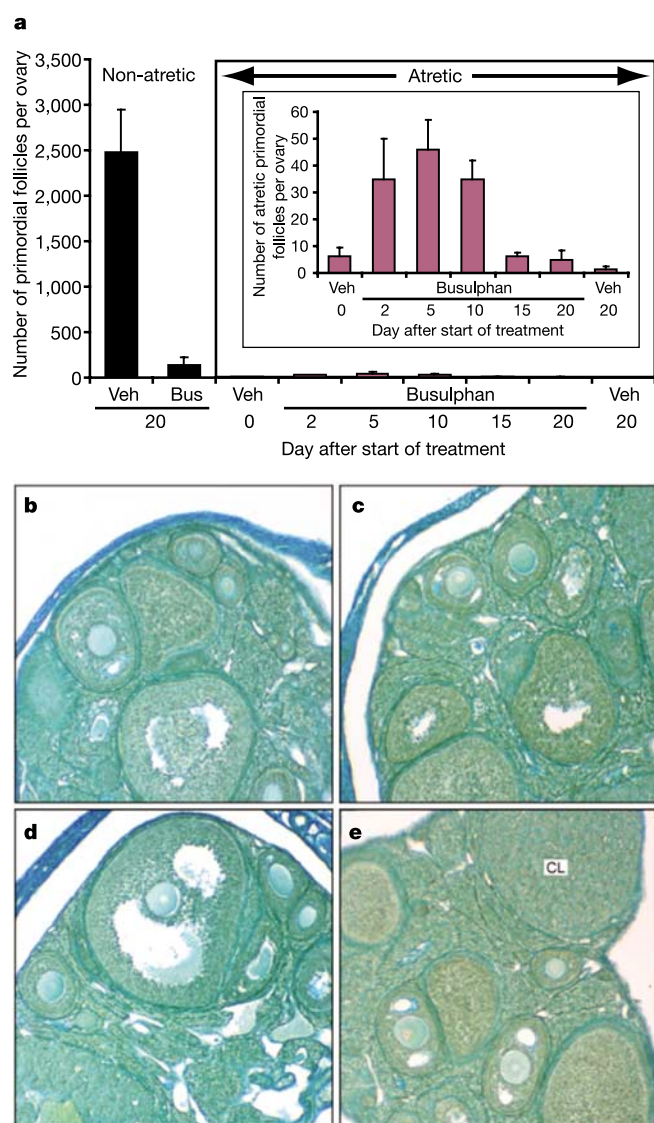


Figure 4 Busulphan eliminates the primordial follicle reserve in adult female mice. **a**, Numbers of non-atretic and atretic primordial follicles present in the ovaries of vehicle (Veh)- or busulphan (Bus)-treated mice (mean $\pm$ standard error, $n = 3$ –4 mice per data point). The inset shows results for primordial follicle atresia. **b–e**, Histological appearance of ovaries of vehicle-treated (**b**) or busulphan-treated (**c–e**) mice, underscoring the lack of nonspecific or widespread cytotoxic effects of busulphan on the ovaries or existing oocytes within developing follicles. Note also the presence of corpora lutea (CL) in busulphan-exposed ovaries (**e**), indicative of normal ovulatory function in these animals.

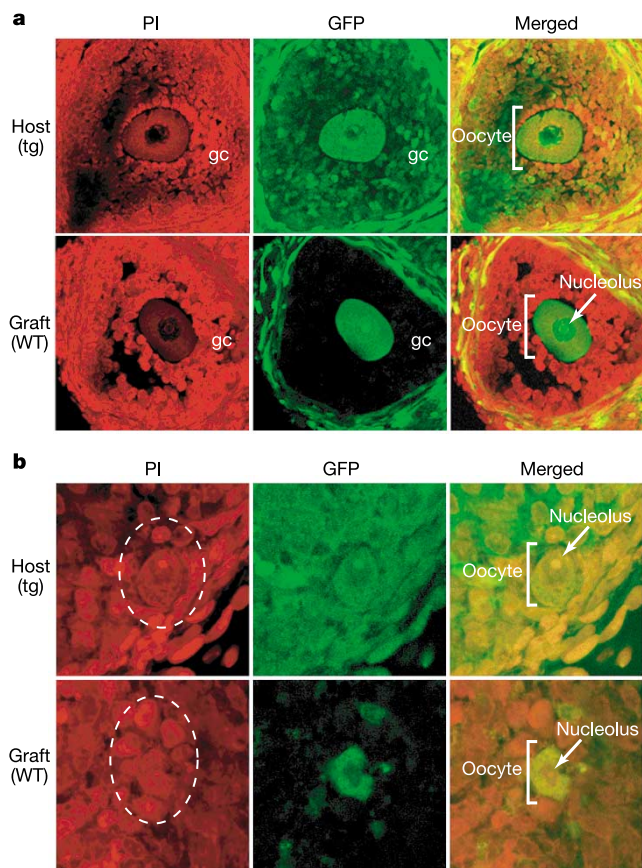


Figure 5 GFP-transgenic germ cells form follicles in wild-type ovaries. **a, b**, GFP expression (green) in sections of host (GFP-transgenic, tg) and grafted (wild type, WT) ovarian tissues counterstained with propidium iodide (PI; red). **a**, Antral follicle in grafted ovarian tissue containing a GFP-positive oocyte enclosed within GFP-negative granulosa cells (gc). A maturing antral follicle in the host ovary is shown for comparison. **b**, Primary follicle in grafted ovarian tissue containing a GFP-positive oocyte enclosed within GFP-negative granulosa cells (broken white line). A primary follicle in the host ovary is shown for comparison. Additional examples and details are provided in the Supplementary Information.

follicle-enclosed, GFP-positive oocytes in the wild-type ovarian fragments that were indistinguishable from follicle-enclosed oocytes in the host ovarian tissue (Fig. 5; see also Supplementary Fig. S3). Moreover, the granulosa cells enveloping the GFP-positive oocytes in the grafts were negative for GFP, indicating that transgenic germ cells had infiltrated the grafted tissue and initiated folliculogenesis with the resident wild-type somatic cells.

This latter finding, when considered with results from past studies of mammalian stem-cell migration to their natural niches after introduction into a host^{41–44}, suggests that GSCs exist in the postnatal mouse ovary. In the testis and in the *Drosophila* ovary, GSCs are defined by their capacity to self-renew after division while simultaneously producing a daughter cell capable of differentiating into a mature egg or sperm—a process referred to as asymmetric division^{13–15}. To sustain the addition of new primordial follicles during juvenile and adult life, the mouse ovary must possess either a small pool of asymmetrically dividing GSCs or a large pool of non-renewing, pre-meiotic germ cells that produce oocytes after symmetric divisions. Given that histological surveys did not reveal any evidence in support of the latter possibility (data not shown), the large ovoid germ cells in the surface epithelium (Fig. 2a, b) were then considered as candidates for GSCs. Histomorphometric studies at day 30 postpartum revealed the presence of 63 ± 8 such cells per ovary (mean $\pm$ standard error, $n = 4$ mice), a number close to that expected for a small pool of asymmetrically dividing germ cells.

Discussion

In 1921, Pearl and Schoppe cited a “basic biological doctrine that during the life of the individual there neither is nor can be any increase in the number of primary oocytes beyond those originally laid down when the ovary was formed”⁴⁵. This concept was solidified as dogma in 1951 in a paper that critically evaluated, and effectively dispelled, any work contrary to the belief that mammalian females are endowed with a finite and non-renewing germ-cell reserve during the perinatal period¹. Although this dogma has persisted for more than 50 yr, the present study provides evidence that challenges the validity of this belief, which represents one of the most basic underpinnings of reproductive biology. The data shown also underscore the significance of mammalian female GSCs to regenerating the follicle reserve in adult life. If one considers that the total pool of non-atretic immature follicles per ovary in C57BL/6 mice is less than 4,000 at puberty, the presence of between 700–1,300 atretic immature follicles at this time point, even with a 3-day clearance rate for their removal, would be predictive of complete exhaustion of the follicle reserve in a few weeks. However, C57BL/6 female mice possess primordial and early growing follicles past 12 months of age¹¹. Therefore, in addition to providing new directions to explore with respect to elucidating the biology of mammalian female GSCs, this work has significant clinical implications related to therapeutic expansion of the follicle reserve as a means to postpone normal or premature ovarian failure. □

Methods

Animals

Age-specified or timed-pregnant wild-type C57BL/6 and CD1 female mice were purchased from Charles River Laboratories, whereas AKR/J mice were obtained from Jackson Laboratories. For the atresia synchronization and oocyte clearance experiments, C57BL/6 female mice were given a single intraperitoneal injection of vehicle (corn oil) or DMBA (80 mg per kg body weight; re-suspended in corn oil) on day 25 postpartum, and ovaries were collected just before injection and at 24 h intervals after injection for up to 96 h. For the busulphan experiments, C57BL/6 female mice were given an intraperitoneal injection of vehicle (DMSO) or busulphan (20 mg per kg body weight; re-suspended in DMSO) on day 25 and again on day 35 postpartum, and ovaries were collected at the indicated times for analysis. Transgenic mice with ubiquitous expression of GFP⁴⁰ were obtained from Jackson Laboratories (strain: STOCK TgN(GFPU)5Nagy) and used for ovarian grafting experiments as described below (see ‘Ovarian grafting’ section of Methods). All animal work was conducted using procedures reviewed and approved by the institutional animal care and use committee of Massachusetts General Hospital, and were

conducted in accordance with the NIH Guide for the Care and Use of Laboratory Research Animals.

Ovarian histology and follicle counts

Ovaries were fixed (0.34 N glacial acetic acid, 10% formalin, 28% ethanol), paraffin embedded, serially sectioned (8 μ m), aligned in order on glass microscope slides, and stained with haematoxylin and picric methyl blue. The number of non-atretic or atretic primordial, primary and preantral follicles was then determined, as described^{9,46} (see Supplementary Information for additional details). Immature follicles were scored as atretic if the oocyte was degenerating (convoluted, condensed) or fragmented. Grossly atretic immature follicles lacking oocyte remnants were not included in the analyses.

Immunohistochemistry

After fixation in 4% neutral-buffered paraformaldehyde and embedding in paraffin, 6- μ m tissue sections were cut from the ovaries and mounted on slides. The sections were de-waxed in xylenes, re-hydrated, and boiled for 5 min in 10 mM sodium citrate using a microwave. Primary antibodies specific for MVH²⁶, BrdU (Zymed), or SCP3 (refs 47, 48) were then used for immunohistochemical analyses as per the suppliers’ recommendations (see Supplementary Information for details). To visualize chromatin morphology within MVH-positive cells, MVH detection in fixed ovarian tissue sections was performed, and the sections were then incubated in 10 μ g ml⁻¹ RNase A for a minimum of 2 h at 37 °C. After washing, the sections were mounted in the presence of 1.5 μ g ml⁻¹ propidium iodide (Vector Labs), as described⁴⁹, and visualized using a Zeiss LSM 5 Pascal confocal microscope. For SCP3 immunohistochemistry, photomicrographs of the sections were taken under Hoffman optics without prior counterstaining to prevent masking of the immunoreaction signal by vital dyes.

Reverse transcription PCR analysis

For ovaries collected at each time point and for control tissues, total RNA was extracted and 1 μ g of total RNA was reverse transcribed (Superscript II RT; Invitrogen) using oligo-dT primers. Amplification via 28 cycles of PCR was performed using *Taq* polymerase and buffer D (Epicentre) with primer sets specific for each gene (Supplementary Table S2). The ribosomal gene *L7* was co-amplified and used as a loading control for each sample, and 28 cycles were found to be within the linear range of amplification for each experimental primer set (data not shown). All PCR products were isolated, subcloned and sequenced for confirmation. In those samples showing more than one amplified product per primer set, each band was isolated, subcloned and sequenced. These additional bands were determined to be known splice variants of the targeted genes (*Dmc-1/Dmc-1d*; *Spo11a/Spo11b*) (Supplementary Table S2).

Ovarian grafting

Heterozygous transgenic male and female mice with ubiquitous expression of GFP⁴⁰ were mated to generate wild-type and transgenic female offspring for intrabursal ovarian grafting. Briefly, young adult (58–69 days postpartum) transgenic female mice were anaesthetized (avertin, 200 mg per kg, intraperitoneal) to expose one of the two ovaries in each mouse through dorso-lateral incisions. For each animal, a small hole was cut in the ovarian bursa laterally near the hilus, and approximately one-half of the host ovary was removed in preparation for grafting. Ovaries collected from donor (wild-type littermate) female mice were bisected, and one-half of a wild-type ovary was placed within the transgenic recipient’s bursal cavity in contact with the remaining host ovarian tissue. The reproductive tract was then allowed to settle back into the peritoneal cavity and the incision was closed. A total of six transgenic hosts were used for this experiment, four of which received unilateral wild-type ovarian grafts while the remaining two received bilateral wild-type ovarian grafts. Between 3–4 weeks after surgery, the ovarian tissues were removed and processed for GFP visualization, after propidium iodide counterstaining, by confocal laser scanning microscopy⁵⁰.

Received 11 September; accepted 23 December 2003; doi:10.1038/nature02316.

1. Zuckerman, S. The number of oocytes in the mature ovary. *Recent Prog. Horm. Res.* **6**, 63–108 (1951).
2. Borum, K. Oogenesis in the mouse. A study of meiotic prophase. *Exp. Cell Res.* **24**, 495–507 (1961).
3. Peters, H. Migration of gonocytes into the mammalian gonad and their differentiation. *Phil. Trans. R. Soc. Lond. B* **259**, 91–101 (1970).
4. McLaren, A. Meiosis and differentiation of mouse germ cells. *Symp. Soc. Exp. Biol.* **38**, 7–23 (1984).
5. Anderson, L. D. & Hirshfield, A. N. An overview of follicular development in the ovary: from embryo to the fertilized ovum *in vitro*. *Md Med. J.* **41**, 614–620 (1992).
6. Faddy, M. J., Jones, E. C. & Edwards, R. G. An analytical model for ovarian follicle dynamics. *J. Exp. Zool.* **197**, 173–186 (1976).
7. Faddy, M. J., Telfer, E. & Gosden, R. G. The kinetics of pre-antral follicle development in ovaries of CBA/Ca mice during the first 14 weeks of life. *Cell Tissue Kinet.* **20**, 551–560 (1987).
8. Faddy, M. J. Follicle dynamics during ovarian ageing. *Mol. Cell. Endocrinol.* **163**, 43–48 (2000).
9. Perez, G. I. *et al.* Prolongation of ovarian lifespan into advanced chronological age by *Bax*-deficiency. *Nature Genet.* **21**, 200–203 (1999).
10. Tilly, J. L. Commuting the death sentence: how oocytes strive to survive. *Nature Rev. Mol. Cell Biol.* **2**, 838–848 (2001).
11. Gosden, R. G., Laing, S. C., Felicio, L. S., Nelson, J. F. & Finch, C. E. Imminent oocyte exhaustion and reduced follicular recruitment mark the transition to acyclicity in aging C57BL/6 mice. *Biol. Reprod.* **28**, 255–260 (1983).
12. Richardson, S. J., Senikas, V. & Nelson, J. F. Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. *J. Clin. Endocrinol. Metab.* **65**, 1231–1237 (1987).
13. Lin, H. The two stem cells in the germline. *Annu. Rev. Genet.* **31**, 455–491 (1997).
14. Spradling, A. H., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98–104 (2001).

15. Lin, H. The stem-cell niche theory: lessons from flies. *Nature Rev. Genet.* **3**, 931–940 (2002).
16. Wyllie, A. H., Kerr, J. F. R. & Currie, A. R. Cell death: the significance of apoptosis. *Int. Rev. Cytol.* **68**, 251–306 (1980).
17. Ijiri, K. & Potten, C. S. Response of intestinal cells of differing topographical and hierarchical status to ten cytotoxic drugs and five sources of radiation. *Br. J. Cancer* **47**, 175–185 (1983).
18. Bursch, W., Paffe, S., Putz, B., Barthel, G. & Schulte-Hermann, R. Determination of the length of the histological stages of apoptosis in normal liver and in altered hepatic foci of rats. *Carcinogenesis* **11**, 847–853 (1990).
19. Ratts, V. S., Flaws, J. A., Kolp, R., Sorenson, C. M. & Tilly, J. L. Ablation of *bcl-2* gene expression decreases the numbers of oocytes and primordial follicles established in the post-natal female mouse gonad. *Endocrinology* **136**, 3665–3668 (1995).
20. Pepling, M. E. & Spradling, A. C. Mouse ovarian germ cell cysts undergo programmed breakdown to form primordial follicles. *Dev. Biol.* **234**, 339–351 (2001).
21. Mattison, D. R. Morphology of oocyte and follicle destruction by polycyclic aromatic hydrocarbons in mice. *Toxicol. Appl. Pharmacol.* **53**, 249–259 (1980).
22. Matikainen, T. et al. Aromatic hydrocarbon receptor-driven *Bax* gene expression is required for premature ovarian failure caused by biohazardous environmental chemicals. *Nature Genet.* **28**, 355–360 (2001).
23. Canning, J., Takai, Y. & Tilly, J. L. Evidence for genetic modifiers of ovarian follicular endowment and development from studies of five inbred mouse strains. *Endocrinology* **144**, 9–12 (2003).
24. Crone, M., Levy, E. & Peters, H. The duration of the premeiotic DNA synthesis in mouse oocytes. *Exp. Cell Res.* **39**, 678–688 (1965).
25. Morita, Y. et al. Requirement for phosphatidylinositol-3'-kinase in cytokine-mediated germ cell survival during fetal oogenesis in the mouse. *Endocrinology* **140**, 941–949 (1999).
26. Fujiwara, Y. et al. Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila* vasa and its specific expression in germ cell lineage. *Proc. Natl Acad. Sci. USA* **91**, 12258–12262 (1994).
27. Noce, T., Okamoto-Ito, S. & Tsunekawa, N. Vasa homolog genes in mammalian germ cell development. *Cell Struct. Funct.* **26**, 131–136 (2001).
28. Selden, J. R. et al. Statistical confirmation that immunofluorescent detection of DNA repair in human fibroblasts by measurement of bromodeoxyuridine incorporation is stoichiometric and sensitive. *Cytometry* **14**, 154–167 (1993).
29. Davis, A. F. & Clayton, D. A. *In situ* localization of mitochondrial DNA replication in intact mammalian cells. *J. Cell Biol.* **135**, 883–893 (1996).
30. Yuan, L. et al. The murine *SCP3* gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Mol. Cell* **5**, 73–83 (2000).
31. Yuan, L. et al. Female germ cell aneuploidy and embryo death in mice lacking the meiosis-specific protein *SCP3*. *Science* **296**, 1115–1118 (2002).
32. Cohen, P. & Pollard, J. W. Regulation of meiotic recombination and prophase I progression in mammals. *Bioessays* **23**, 996–1009 (2001).
33. Gosden, R., Clarke, H. & Miller, D. in *Reproductive Medicine—Molecular, Cellular and Genetic Fundamentals* (ed. Fauser, B. C. J. M.) 365–380 (Parthenon, New York, 2003).
34. Bucci, L. R. & Meistrich, M. L. Effects of busulfan on murine spermatogenesis: cytotoxicity, sterility, sperm abnormalities, and dominant lethal mutations. *Mutat. Res.* **176**, 259–268 (1987).
35. Brinster, R. L. & Zimmermann, J. W. Spermatogenesis following male germ-cell transplantation. *Proc. Natl Acad. Sci. USA* **91**, 11298–11302 (1994).
36. Brinster, C. J. et al. Restoration of fertility by germ cell transplantation requires effective recipient preparation. *Biol. Reprod.* **69**, 412–420 (2003).
37. Pelloux, M. C., Picon, R., Gangnerau, M. N. & Darmoul, D. Effects of busulphan on ovarian folliculogenesis, steroidogenesis and anti-Müllerian activity of rat neonates. *Acta Endocrinol.* **118**, 218–226 (1988).
38. Generoso, W. M., Stout, S. K. & Huff, S. W. Effects of alkylating agents on reproductive capacity of adult female mice. *Mutat. Res.* **13**, 171–184 (1971).
39. Shiromizu, K., Thorgeirsson, S. S. & Mattison, D. R. Effect of cyclophosphamide on oocyte and follicle number in Sprague–Dawley rats, C57BL/6N and DBA/2N mice. *Pediatr. Pharmacol.* **4**, 213–221 (1984).
40. Hadjantonakis, A. K., Gertsenstein, M., Ikawa, M., Okabe, M. & Nagy, A. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech. Dev.* **76**, 79–90 (1998).
41. Nagano, M. C. Homing efficiency and proliferation kinetics of male germ line stem cells following transplantation in mice. *Biol. Reprod.* **69**, 701–707 (2003).
42. Szilvassy, S. J., Ragland, P. L., Miller, C. L. & Eaves, C. J. The marrow homing efficiency of murine hematopoietic stem cells remains constant during ontogeny. *Exp. Hematol.* **31**, 331–338 (2003).
43. Torrente, Y. et al. Identification of a putative pathway for the muscle homing of stem cells in a muscular dystrophy model. *J. Cell Biol.* **162**, 511–520 (2003).
44. Oh, H. et al. Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction. *Proc. Natl Acad. Sci. USA* **100**, 12313–12318 (2003).
45. Pearl, R. & Schoppe, W. F. Studies on the physiology of reproduction in the domestic fowl. *J. Exp. Zool.* **34**, 101–118 (1921).
46. Tilly, J. L. Ovarian follicle counts—not as simple as 1, 2, 3. *Reprod. Biol. Endocrinol.* **1**, 11 (2003).
47. Walpita, D., Plug, A. W., Neff, N. F., German, J. & Ashley, T. Bloom's syndrome protein, BLM, colocalizes with replication protein A in meiotic prophase nuclei of mammalian spermatocytes. *Proc. Natl Acad. Sci. USA* **96**, 5622–5627 (1999).
48. Russell, L. B., Hunsicker, P. R., Hack, A. M. & Ashley, T. Effect of the topoisomerase-II inhibitor etoposide on meiotic recombination in male mice. *Mutat. Res.* **464**, 201–212 (2000).
49. Foley, J. G. D. & Bard, J. B. L. Apoptosis in the cortex of the developing mouse kidney. *J. Anat.* **201**, 477–484 (2002).
50. Walter, I. et al. Rapid and sensitive detection of enhanced green fluorescent protein expression in paraffin sections by confocal laser scanning microscopy. *Histochem. J.* **32**, 99–103 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Y. Morita for technical assistance; T. Noce for MVH antiserum; T. Ashley for SCP3 antiserum; and I. Schiff and F. Frigoletto Jr for critical reading of the manuscript before its submission. This work was supported by the National Institute on Aging and by Vincent Memorial Research Funds. This study was conducted while J.L.T. was an Investigator of the Steven and Michele Kirsch Foundation, and while J.J. was a Research Fellow supported by The Lalor Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L.T. (jtilly@partners.org).

Convective-region geometry as the cause of Uranus' and Neptune's unusual magnetic fields

Sabine Stanley & Jeremy Bloxham

Department of Earth and Planetary Sciences, Harvard University,
20 Oxford Street, Cambridge, Massachusetts 02138, USA

The discovery of Uranus' and Neptune's non-dipolar, non-axisymmetric magnetic fields^{1–4} destroyed the picture—established by Earth, Jupiter and Saturn^{5–6}—that planetary magnetic fields are dominated by axial dipoles. Although various explanations for these unusual fields have been proposed^{3,7–10}, the cause of such field morphologies remains unexplained. Planetary magnetic fields are generated by complex fluid motions in electrically conducting regions of the planets (a process known as dynamo action), and so are intimately linked to the structure and evolution of planetary interiors. Determining why Uranus and Neptune have different field morphologies is not only critical for studying the interiors of these planets, but also essential for understanding the dynamics of magnetic-field generation in all planets. Here we present three-dimensional numerical dynamo simulations that model the dynamo source region as a convecting thin shell surrounding a stably stratified fluid interior. We show that this convective-region geometry produces magnetic fields similar in morphology to those of Uranus and Neptune. The fields are non-dipolar and non-axisymmetric, and result from a combination of the stable fluid's response to electromagnetic stress and the small length scales imposed by the thin shell.

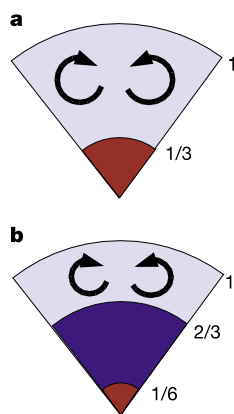


Figure 1 Dynamo model geometries. Interior planetary geometries used in numerical modelling of Earth's dynamo (a) and Uranus' and Neptune's dynamos (b). A radius measure of 1 corresponds to the top of the dynamo source region. Solid regions are in brown, convective fluid regions are in light blue and stably stratified fluid regions are in dark blue. In our numerical dynamo models for Uranus and Neptune, we make the conductivity of the stable shell equal to that of the convecting shell so that the only difference between the two layers is their stability to convection. The method of maintaining stability and instability is similar to that of refs 27 (a study of thermal convection in the presence of stable layers) and 28 (in which a thin stable layer near the core–mantle boundary was incorporated in a 2.5D geodynamo model), although our actual stability profiles are different. In the convectively unstable shell, we impose a super-adiabatic background temperature gradient consistent with our fixed heat flux boundary conditions: $\partial T_0/\partial r \propto -r^{-2}$. In our stable shell we impose a sub-adiabatic temperature gradient: $\partial T_0/\partial r \propto c$, where c is a positive constant.

Uranus' and Neptune's magnetic fields are probably generated in their ionically conducting, fluid 'ice' (H_2O , CH_4 , NH_3 and H_2S , not necessarily in the form of intact molecules¹¹) layers, which extend out to about 0.75 and 0.8 of their total radii, respectively¹². Podolak *et al.*⁸ and Hubbard *et al.*⁹ examined thermal evolution models that satisfy the present-day observed luminosities of Uranus and Neptune, and found that the interior portion of these planets' ice layers may be compositionally stably stratified and therefore unable to convect. They proposed that a subtle difference in the stratification between the two planets could explain the seemingly contradictory observation that the two planets have very similar internal composition and structure, yet very different internal heat flows. They also suggested that this geometry of a thin shell dynamo surrounding a stably stratified fluid interior might explain the non-dipolar, non-axisymmetric magnetic fields of Uranus and Neptune.

Here we use the numerical dynamo model of refs 13 and 14 to test whether this geometry is favourable for producing non-dipolar, non-axisymmetric magnetic fields. Early models of the Earth's dynamo were successful at reproducing many of the salient features of its magnetic field, including westward drift, reversals and most notably, the axial dipole dominance of the field^{13,15,16}. More recent numerical models (see ref. 17 for a review) that implement different numerical methods or work in different parameter regimes still produce predominantly axially dipolar fields. A handful of non-axial, non-dipolar dynamos have been found^{18–22} when working in specific regions of parameter space, but because all numerical models at present cannot operate in the appropriate parameter space for planetary cores, we seek solutions that prevail over a wide range of parameter values. If non-dipolar, non-axisymmetric solutions result from changing only the geometry of the problem, then it would seem a more likely and more robust explanation for Uranus' and Neptune's magnetic fields.

In geodynamo models, the implemented geometry is consistent with the Earth's internal structure, in which the magnetic field is generated by convection in the fluid-iron outer core of the planet surrounding the small solid-iron inner core (Fig. 1a). In order to study Uranus' and Neptune's magnetic fields, we implement the geometry proposed by Podolak *et al.*⁸ and Hubbard *et al.*⁹ in which the magnetic field is generated by convection in a thin shell surrounding a stably stratified fluid core (Fig. 1b). Table 1 lists the control parameters for our model.

We find that this geometry produces fields similar in morphology

Table 1 Control parameters of our dynamo model

Parameter	Value
r_{io}	1/6
r_{so}	2/3
$E = \frac{\nu}{2\Omega d^2}$	2×10^{-5}
$Ro = \frac{\eta}{2\Omega d^2}$	2×10^{-5}
$q_k = \frac{\kappa}{\eta}$	1
$Ra = \frac{\alpha_T g_0 h_T d^2}{2\Omega \eta}$	18,000–24,000

Definitions and values of control parameters in our numerical model. The parameters given in the table are the ratio of the solid inner core to total core radius (r_{io}), the ratio of the stable fluid shell to total core radius (r_{so}), and the Ekman, magnetic Rossby, Prandtl and Rayleigh numbers (E , Ro , q_k and Ra , respectively). These non-dimensional numbers depend on the kinematic viscosity (ν), magnetic and thermal diffusivities (η and κ , respectively), core rotation rate (Ω), outer core thickness (d), thermal expansion coefficient (α_T), gravity (g_0) and the characteristic heat flux (h_T). The small solid inner core in our models is a numerical necessity, but may also be physically realistic because Uranus and Neptune probably contain small rocky solid cores²⁶. Except for the core radius and stability, all other parameters are the same as those in Kuang and Bloxham geodynamo models. As well as the prescribed parameters in the table, we use impenetrable, stress-free boundary conditions on the velocity and fixed heat flux boundary conditions on the temperature. We use the same hyperdiffusivities as ref. 14 on the thermal, viscous and magnetic diffusivities, but we do not believe that our non-dipolar, non-axial solutions result from the use of hyperdiffusivities, because current geodynamo models using similar hyperdiffusivities produce axially dipolar dominated solutions.

to those of Uranus and Neptune. In Fig. 2, substantial non-dipolar, non-axisymmetric structure is evident in the surface radial magnetic field of our model. This model is highly variable in time, so one snapshot is not indicative of the field at other times, but can qualitatively capture the similarity in complexity of our numerical model and Uranus' and Neptune's observed fields, as well as the difference between our model and Earth-like fields.

To quantify the degree of symmetry and dipole dominance in our model, we examine the spectral power in the magnetic field. In contrast to the Earth's magnetic power spectrum, which is dominated by the dipole component at the Earth's surface (it is at least an order of magnitude larger than the quadrupole and octupole components), Uranus' and Neptune's spectra have comparable power in all three components, as does our numerical model (Fig. 3a). The Earth's field is also dominated by axisymmetric components unlike the observed fields of Uranus and Neptune, which have significant power in both axisymmetric and non-axisymmetric terms (Fig. 3b). A prominent feature in Uranus' and Neptune's spectra is the peak in power at spherical harmonic order $m = 1$. Our model also has significant power in axisymmetric and non-axisymmetric modes, and also produces this peak at $m = 1$.

By altering two properties of the planetary geometry, namely the thickness of the outer convecting shell, and the state (fluid versus solid) of the inner core, we have obtained magnetic-field morphologies very different from Earth-like models. We have also examined whether changing only the thickness of the convecting shell, while keeping the inner core solid, would be enough to produce this new field morphology. We find that dynamo models operating in a thin shell surrounding a large solid inner core with the same conductivity as the outer core are axially dipolar dominated.

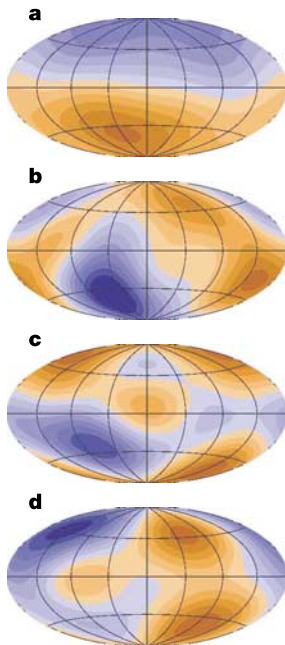


Figure 2 Surface radial magnetic fields. The radial component of the magnetic field from spherical harmonic models truncated at degree and order 3 at the planetary surfaces of Earth (a), Uranus (b), Neptune (c) and our numerical model (d) are shown. For Uranus and Neptune, the surface is defined as the 1-bar pressure level. Gauss coefficients are from ref. 29 for Earth, and ref. 30 for Uranus and Neptune. For the numerical model, the dynamo source region is assumed to be at 0.75 planetary radii. The colour contours extend from +0.1 mT (dark orange) to -0.1 mT (dark blue).

These results are consistent with those of ref. 23, in which the potential future geodynamo operating in a thin shell geometry was examined.

The stabilizing effects that a solid conducting inner core can have on a dynamo were first demonstrated in ref. 24. Although it has been shown²⁵ that the Earth's relatively small inner core (inner core radius of ~1,220 km occupies ~4% of the total core volume) may have little effect on the geodynamo, we believe the much larger relative size of the inner core in a thin shell dynamo (inner core radius ~2/3 of the total core radius occupies ~30% of the total core volume) does affect the field morphology. As the conducting inner core is in solid-body rotation, the strong differential rotation between the magnetically coupled inner and outer cores acts to make the field

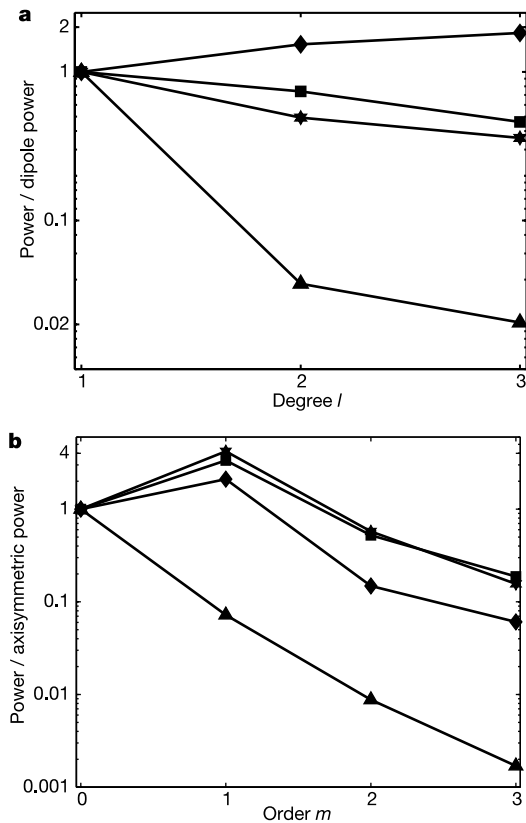


Figure 3 Surface magnetic power spectra. Outside the dynamo generation region, the magnetic field can be represented by the gradient of a scalar potential ($\mathbf{B} = -\nabla \Phi$) which is expanded in spherical harmonics as

$$\Phi = a \sum_{l=1}^{\infty} \left(\frac{a}{r} \right)^{l+1} P_l^m(\cos \theta) [g_l^m \cos(m\phi) + h_l^m \sin(m\phi)]$$

where a is the planetary radius, (r, θ, ϕ) are spherical coordinates and P_l^m are Schmidt-normalized associated Legendre polynomials. We define the power as the mean square field intensity which is given in terms of the Gauss coefficients g_l^m and h_l^m for each harmonic degree l and order m as:

$$p(l, m, r) = (l+1) \left(\frac{a}{r} \right)^{2(l+4)} [(g_l^m)^2 + (h_l^m)^2]$$

The spectra at the surface of Earth (triangles), Uranus (squares), Neptune (diamonds) and our numerical model (stars) are shown. For the numerical model, the average power over a magnetic diffusion time is shown. Power versus degree l is found by summing over the order m components for each degree l (a). The dipole, quadrupole and octupole power are given by the $l = 1, 2$ and 3 components, respectively. Power versus order is found by summing over the degree l components for each order m (b). The $m = 0$ term is the axisymmetric power, and the higher-order m terms represent the non-axisymmetric power.

more axisymmetric. Also, the magnetic field that penetrates the conducting solid inner core can only vary on the much longer magnetic diffusion timescale rather than the shorter timescale of fluid motions in the outer core. This effectively anchors the magnetic field and stabilizes it against the rapid fluctuations in the outer core, allowing the field to maintain a more stable dipole.

If the region interior to the convecting shell is fluid instead of solid, then it can respond to electromagnetic stress through fluid motion and is not constrained by the stabilizing effects of a solid conducting core. As the fluid inner region is not in solid-body rotation, the axisymmetrizing effect of strong differential rotation is suppressed. Also, the magnetic field that penetrates the inner region can vary by moving the fluid instead of only through magnetic diffusion. This reduction in stability, combined with the smaller length scales of the thin shell, promote the creation of higher-degree, non-axisymmetric structure in our numerical models with stably stratified fluid interiors.

As the stabilizing effects of the inner core are a result of its solid state and conductivity, our theory predicts that a dynamo operating in a thin shell surrounding an electrically insulating solid inner core would also produce non-dipolar, non-axisymmetric fields. We have created numerical dynamos operating in a thin shell (similar radius to our convecting shell surrounding the stably stratified interior) surrounding a large solid inner core with a much smaller electrical conductivity than the outer core (inner to outer core conductivity ratios of 0.1–0.0001), and found that they do produce non-dipolar, non-axisymmetric magnetic fields, as predicted.

Future missions to Uranus and Neptune to study further these planets' magnetic fields and internal properties could determine the validity of our model. More accurate surface power spectra extending out to higher degree and order could be produced with magnetic-field measurements acquired closer to the planets with more global spatial distribution. These spectra could then be compared to our numerical model to assess the model's validity. The evolution in time of the magnetic fields could also be compared to our numerical model by examining the changes in the fields since the Voyager II observations. Although no missions to Uranus and Neptune are currently planned, the upcoming data from MESSENGER and Cassini (on the magnetic fields of Mercury and Saturn, respectively) will provide tests of the implication of our results: convective-region geometry is an important factor in determining magnetic-field morphology, albeit in a different context from this study. □

Received 9 October 2003; accepted 28 January 2004; doi:10.1038/nature02376.

- Ness, N. F. *et al.* Magnetic fields at Uranus. *Science* **233**, 85–89 (1986).
- Ness, N. F. *et al.* Magnetic fields at Neptune. *Science* **246**, 1473–1478 (1989).
- Connerney, J. E. P., Acuna, M. H. & Ness, N. F. The magnetic field of Uranus. *J. Geophys. Res.* **92**, 15329–15336 (1987).
- Connerney, J. E. P., Acuna, M. H. & Ness, N. F. The magnetic field of Neptune. *J. Geophys. Res.* **96**, 19023–19042 (1991).
- Connerney, J. E. P. Magnetic fields of the outer planets. *J. Geophys. Res.* **98**, 18659–18679 (1993).
- Russell, C. T. Magnetic fields of the terrestrial planets. *J. Geophys. Res.* **98**, 18681–18695 (1993).
- Ruzmaikin, A. A. & Starchenko, S. V. On the origin of Uranus and Neptune magnetic fields. *Icarus* **93**, 82–87 (1991).
- Podolak, M., Hubbard, W. B. & Stevenson, D. J. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 29–61 (Univ. Arizona Press, Tucson, 1991).
- Hubbard, W. B., Podolak, M. & Stevenson, D. J. in *Neptune and Triton* (ed. Cruikshank, D. P.) 109–138 (Univ. Arizona Press, Tucson, 1995).
- Holme, R. Three-dimensional kinematic dynamos with equatorial symmetry: application to the magnetic fields of Uranus and Neptune. *Phys. Earth Planet. Inter.* **102**, 105–122 (1997).
- Hubbard, W. B. *et al.* Interior structure of Neptune—comparison with Uranus. *Science* **253**, 648–651 (1991).
- Nellis, W. J. *et al.* The nature of the interior of Uranus based on studies of planetary ices at high dynamic pressure. *Science* **240**, 779–781 (1988).
- Kuang, W. & Bloxham, J. An earth-like numerical dynamo model. *Nature* **389**, 371–374 (1997).
- Kuang, W. J. & Bloxham, J. Numerical modelling of magnetohydrodynamic convection in a rapidly rotating spherical shell: weak and strong field dynamo action. *J. Comput. Phys.* **153**, 51–81 (1999).
- Glatzmaier, G. A. & Roberts, P. H. A 3-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle. *Phys. Earth Planet. Inter.* **91**, 63–75 (1995).
- Glatzmaier, G. A. & Roberts, P. H. A 3-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* **377**, 203–209 (1995).

- Dormy, E., Valet, J. P. & Courtillot, V. Numerical models of the geodynamo and observational constraints. *Geochim. Geophys. Geosyst.* **1**, 2000GC000062 (2000).
- Grote, E., Busse, F. H. & Tilgner, A. Regular and chaotic spherical dynamos. *Phys. Earth Planet. Inter.* **117**, 259–272 (2000).
- Grote, E. & Busse, F. H. Hemispherical dynamos generated by convection in rotating spherical shells. *Phys. Rev. E* **62**, 4457–4460 (2000).
- Kutzner, C. & Christensen, U. R. From stable dipolar towards reversing numerical dynamos. *Phys. Earth Planet. Inter.* **131**, 29–45 (2002).
- Ishihara, N. & Kida, S. Equatorial magnetic dipole field intensification by convection vortices in a rotating spherical shell. *Fluid Dyn. Res.* **31**, 253–274 (2002).
- Aubert, J. & Wicht, J. Axial versus equatorial dipolar dynamo models with implications for planetary magnetic fields. *Earth Planet. Sci. Lett.* (in the press).
- Roberts, P. H. & Glatzmaier, G. A. The geodynamo, past, present and future. *Geophys. Astrophys. Fluid Dynam.* **94**, 47–84 (2001).
- Hollerbach, R. & Jones, C. A. Influence of the Earth's inner core on geomagnetic fluctuations and reversals. *Nature* **365**, 541–543 (1993).
- Wicht, J. Inner core conductivity in numerical dynamo simulations. *Phys. Earth Planet. Inter.* **132**, 281–302 (2002).
- Marley, M., Gomez, P. & Podolak, M. Monte Carlo interior models for Uranus and Neptune. *J. Geophys. Res.* **100**, 23349–23353 (1995).
- Zhang, K. K. & Schubert, G. Teleconvection: remotely driven thermal convection in rotating stratified spherical layers. *Science* **290**, 1944–1947 (2000).
- Jones, C. A., Longbottom, A. W. & Hollerbach, R. A self-consistent convection driven geodynamo model, using a mean field approximation. *Phys. Earth Planet. Inter.* **92**, 119–141 (1995).
- Merrill, R. T., McElhinny, M. W. & McFadden, P. L. *The Magnetic Field of the Earth* 31 (Academic, San Diego, 1996).
- Holme, R. & Bloxham, J. The magnetic fields of Uranus and Neptune: methods and models. *J. Geophys. Res.* **101**, 2177–2200 (1996).

Acknowledgements We thank D. Stevenson for comments on the manuscript. This work was supported by NSERC and NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.S. (stanley@geophysics.harvard.edu).

Observation of entanglement between a single trapped atom and a single photon

B. B. Blinov, D. L. Moehring, L.-M. Duan & C. Monroe

FOCUS Center and University of Michigan Department of Physics, Ann Arbor, Michigan 48109-1120, USA

An outstanding goal in quantum information science is the faithful mapping of quantum information between a stable quantum memory and a reliable quantum communication channel¹. This would allow, for example, quantum communication over remote distances², quantum teleportation³ of matter and distributed quantum computing over a 'quantum internet'. Because quantum states cannot in general be copied, quantum information can only be distributed in these and other applications by entangling the quantum memory with the communication channel. Here we report quantum entanglement between an ideal quantum memory—represented by a single trapped ¹¹¹Cd⁺ ion—and an ideal quantum communication channel, provided by a single photon that is emitted spontaneously from the ion. Appropriate coincidence measurements between the quantum states of the photon polarization and the trapped ion memory are used to verify their entanglement directly. Our direct observation of entanglement between stationary and 'flying' qubits⁴ is accomplished without using cavity quantum electrodynamic techniques^{5–7} or prepared non-classical light sources⁸. We envision that this source of entanglement may be

used for a variety of quantum communication protocols^{2,8} and for seeding large-scale entangled states of trapped ion qubits for scalable quantum computing⁹.

Atom–photon entanglement has been implicit in many previous experimental systems, from early measurements of Bell inequality violations in atomic cascade systems^{10,11} to fluorescence studies in trapped atomic ions^{12,13}. A prime example of current interest is strongly coupled cavity quantum electrodynamics, where individual atoms interact with photons in single-mode cavities^{5–7}. Another example is the continuous-variable entanglement between ensembles of atoms and light fields observed in systems containing macroscopic numbers of atoms and photons^{14–17}. However, atom–photon entanglement has not been directly observed in previous experiments, as the individual atoms and photons have not been under sufficient control. Here we report the explicit demonstration of qubit entanglement between an individual stationary atom and a ‘flying’ optical single photon.

In the experiment, a photon is spontaneously emitted from a single trapped atomic $^{111}\text{Cd}^+$ ion, which is initially excited to a state that has multiple decay channels. Along a certain emission direction selected by an aperture, the photon’s polarization is entangled with particular hyperfine ground states in the de-excited atom. The entanglement is directly verified through subsequent polarization analysis of the photon and state detection of the trapped ion. Given the small acceptance angle of the aperture and other losses, the probability P of detecting a photon in a given trial is small. This is

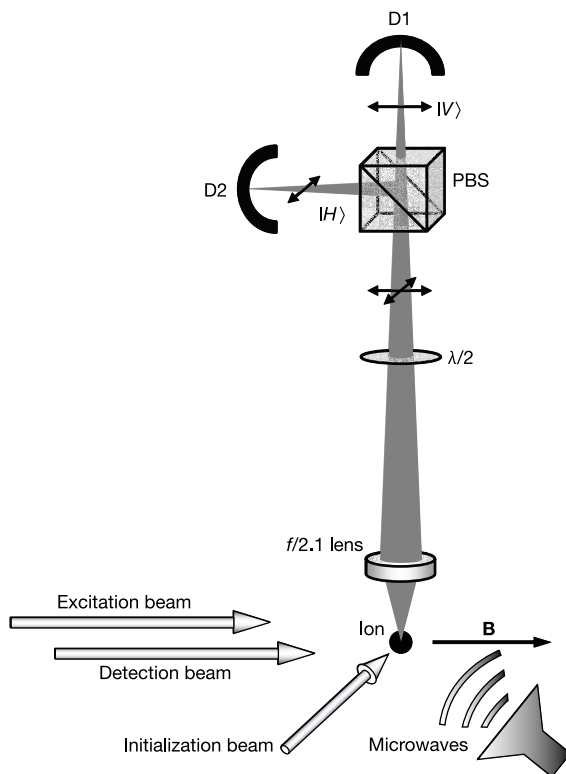


Figure 1 The experimental apparatus. The π -polarized initialization beam propagates perpendicular to the quantization axis defined by a magnetic field $\mathbf{B}$, while the σ^+ -polarized excitation and detection beams propagate along the quantization axis. The scattered photons are collected by an imaging lens and directed to a polarizing beamsplitter (PBS), and two photon-counting detectors D1 and D2 register the V- and H-polarized photons, respectively. The $\lambda/2$ -waveplate is used to rotate the photon polarization for photonic qubit measurements in different bases. A microwave horn is located near the trap to drive coherent transitions between the hyperfine levels of the atomic ground state, at a frequency near 14.5 GHz.

reminiscent of the production of entangled photon pairs through spontaneous optical parametric down-conversion³, but in the current system one of the two daughter qubits resides within a trapped atomic ion—perhaps the most reliable of all qubit memories¹⁸. When accompanied by conventional quantum gates between local trapped ions^{19–21}, this probabilistic entanglement protocol can form the basis for a scalable architecture for quantum communication² and computation⁹.

A diagram of the experimental apparatus is shown in Fig. 1. We trap and laser cool a single $^{111}\text{Cd}^+$ ion in an asymmetric-quadrupole radio frequency trap with a characteristic size of about 0.7 mm. (In this experiment, it is not necessary to laser cool the ion to the Lamb–Dicke limit.) Several laser pulses tuned near the $^2S_{1/2}$ – $^2P_{3/2}$ atomic resonance at 214.5 nm (1) initialize the internal atomic qubit state, (2) excite the atom for the subsequent spontaneous emission of a photon, and (3) detect the internal state of the atom. An applied magnetic field of $B \approx 0.7$ G provides a quantization axis for definition of the photon polarization and the internal atomic qubit levels, stored in $^2S_{1/2}$ hyperfine ground states denoted by quantum numbers $|F, m_F\rangle$, where F is the total angular momentum and m_F is its projection along the quantization axis.

Figure 2 shows a diagram of the relevant energy levels of a $^{111}\text{Cd}^+$ atomic ion, along with the step-by-step description of the experimental procedure. A short π -polarized laser pulse followed by a resonant microwave transfer pulse initializes the ion in the $|1,0\rangle$ state (Fig. 2a). A 50-ns pulse of σ^+ -polarized laser light weakly excites the ion to the $^2P_{3/2}$ $|2,1\rangle$ state, which has a radiative lifetime of $\tau_e \approx 3$ ns. The ion then spontaneously decays to either the $^2S_{1/2}$ $|1,1\rangle$ ground state (defined as $|\downarrow\rangle$) while emitting a π -polarized photon, or the $^2S_{1/2}$ $|1,0\rangle$ ground state (defined as $|\uparrow\rangle$) while emitting a σ^+ -polarized photon (Fig. 2b). The single photon pulses are collected with an $f/2.1$ imaging lens whose axis is perpendicular to the quantization axis. Along this direction, the states of polarization of the σ^+ and the π photons are orthogonal: the former (defined as $|H\rangle$) is polarized perpendicular to the quantization axis,

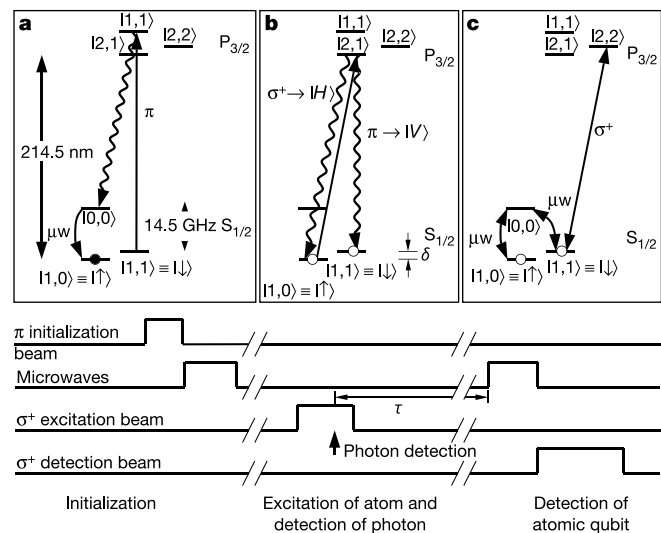


Figure 2 The experimental procedure (time axis not to scale). **a**, The atomic qubit is initialized to the $|\uparrow\rangle$ hyperfine ground state, following a 30- μs π -polarized optical pumping pulse and a 15- μs microwave (μW) rotation. **b**, The atom is weakly excited with a 50-ns σ^+ -polarized optical pulse, resulting in spontaneous emission to either state $|\uparrow\rangle$ or state $|\downarrow\rangle$ (separated by frequency $\delta \approx 2\pi(1.0$ MHz), accompanied by emission and detection of a photon polarized in state $|H\rangle$ or state $|V\rangle$, respectively. **c**, After a delay of $\tau \approx 1$ μs , a 15- μs microwave rotation pulse prepares the atomic qubit for measurement in any basis, and finally a 200- μs σ^+ -polarized optical detection pulse provides efficient measurement of the atomic qubit.

while the latter (defined as $|V\rangle$) is polarized along the quantization axis, as described in the Methods section. The collected light passes through a polarization rotator (a $\lambda/2$ waveplate) followed by a polarizing beamsplitter. The two polarization components are then directed to photon-counting photomultiplier tubes (PMTs), each of quantum efficiency $\eta \approx 20\%$. Following a single photon detection on either PMT, a microwave rotation is applied to the atom, which prepares the atomic qubit to be measured in any basis. This measurement is performed with a σ^+ -polarized detection pulse following standard trapped ion fluorescence techniques²² (Fig. 2c), with an atomic qubit detection efficiency greater than 95%.

The atomic and photonic qubits are entangled following the spontaneous emission because there are two decay channels from the $^2P_{3/2}$ $|2,1\rangle$ excited state, each resulting in distinct photon polarizations. In the experiment, these decay channels are equally likely, but the entanglement is not perfect, as the intensity of the radiation patterns for the π -polarized photons and for the σ^+ -polarized photons differ by a factor of two along the viewing axis. Thus, the resulting state created upon a photon emission in this direction is $\sqrt{3}|H\rangle + \sqrt{3}|V\rangle$, which exhibits an entanglement fidelity of 0.97 under ideal conditions. This entanglement fidelity can (in principle) be made perfect by simply inserting a polarization-selective lossy element into the path of the photons, at the cost of a somewhat lower efficiency²³. Unit fidelity can also be achieved by detecting the scattered photons along the quantization axis, with the initial atom excitation to a $|m_F = 0\rangle$ excited state, such that only σ^+ - and σ^- -polarized photons are collected. We avoid this option, however, as it would require placement of PMTs in a direct line-of-sight with the atomic qubit detection laser beam.

We first measure the conditional probabilities of detecting a certain atomic qubit state given the photonic qubit state after $\sim 1,000$ successful trials. These probabilities are plotted in Fig. 3, with $P(\uparrow|H) = 0.97 \pm 0.01$, $P(\downarrow|H) = 0.03 \pm 0.01$, $P(\uparrow|V) = 0.06 \pm 0.01$ and $P(\downarrow|V) = 0.94 \pm 0.01$, where the errors are statistical. To verify entanglement, we repeat the correlation measurement in a different basis of both photonic and atomic qubits. The photon polarization is rotated by 45° using the $\lambda/2$ waveplate, and the atomic qubit is rotated by applying microwaves driving the $|10\rangle\text{--}|00\rangle$ and the $|11\rangle\text{--}|00\rangle$ transitions as indicated in Fig. 2c. Both qubit rotations are through a Bloch polar angle of $\pi/2$, and the relative phase of the photonic versus atomic qubit rotations is given by $\phi = \delta\tau + \phi_\mu$, where $\delta \approx 2\pi(1.0\text{ MHz})$ is the Zeeman splitting between the $|1,0\rangle$ and the $|1,1\rangle$ levels, and $\tau \approx 1\text{ }\mu\text{s}$ is the

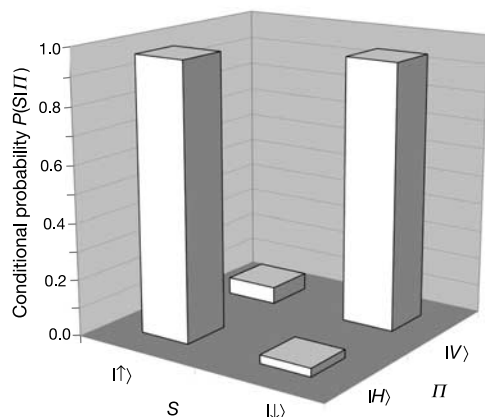


Figure 3 Measured conditional probabilities in the original basis (no atomic or photonic qubit rotation before measurement). The bars indicate the probabilities $P(S|II)$ of detecting atomic qubit states $S = \uparrow$ or $\downarrow$ conditioned upon detecting photon qubit states $II = H$ or V .

time delay between the photon emission and the application of the microwave pulse of phase ϕ_μ . Varying the relative phase of the two qubit rotations by adjusting ϕ_μ produces the correlation fringes shown in Fig. 4a. Figure 4b shows the values of conditional probabilities at the point of maximum correlation; these probabilities are $P(\uparrow|H) = 0.89 \pm 0.01$, $P(\downarrow|H) = 0.11 \pm 0.01$, $P(\uparrow|V) = 0.06 \pm 0.01$ and $P(\downarrow|V) = 0.94 \pm 0.01$. If the atomic and photonic qubits were not entangled but instead prepared in a statistical mixture, then all of these conditional probabilities would have been 0.5.

From these measured correlations, we calculate a bound on the entanglement fidelity to be $F \geq 0.87$ (described in the Methods section), somewhat lower than the potential 0.97 fidelity described above, but still significantly larger than the entanglement threshold of $F > 0.5$ (ref. 24). Several factors contribute to this decrease in fidelity, including: multiple excitations of the atom during the pump pulse (2.5%), mixing of the photon polarizations owing to the nonzero solid angle (0.5%), imperfect rotations of the atomic qubit, mainly due to a 50-ns jitter in the delay τ (1.5%), background counts and dark counts on the PMTs leading to false positives (5–10%), and imperfections in the polarizing beamsplitter (3%). We estimate that magnetic field fluctuations affecting the atomic qubit reduce the fidelity by $\ll 1\%$. All sources of errors combine to give a $\sim 9\%$ reduction of entanglement fidelity, consistent with the observation.

The success probability P of creating an entangled state in a given trial depends on the efficiency of generating a single photon and

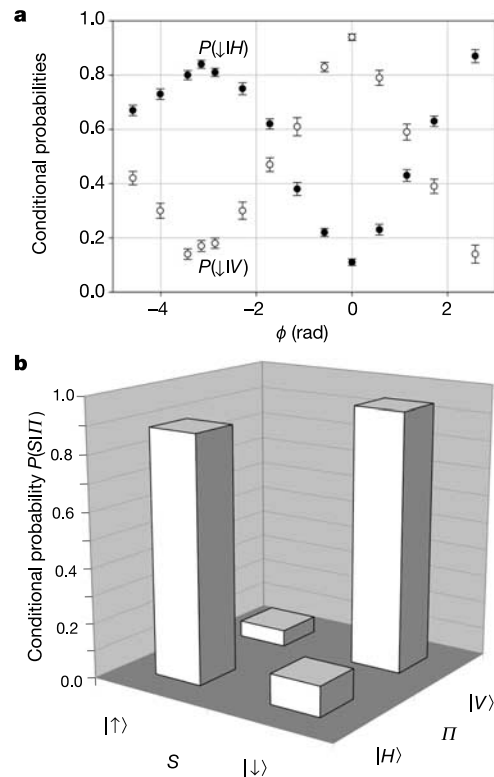


Figure 4 Conditional probabilities after both atomic and photonic qubits are rotated by a polar angle of $\pi/2$ in the Bloch sphere. **a**, Measured conditional probabilities $P(\downarrow|IH)$ and $P(\downarrow|IV)$ as the relative phase ϕ between the atomic and photonic rotations is varied. **b**, Measured conditional probabilities $P(S|II)$ at the point of highest correlation, defined as $\phi = 0$. If the atomic and photonic qubits were not entangled but instead prepared in a statistically mixed state, then all conditional probabilities in the figure would have been 0.5.

then detecting the emitted photon. In our excitation scheme the probability of emitting a single photon in each trial is $P_{\text{exc}} \approx 0.1$ to suppress the multiple-excitation rate. The efficiency of a single photon detection in turn depends on the light collection solid angle $\Delta\Omega$, the transmission efficiency T of the optical elements, and the quantum efficiency of the detectors themselves. The success probability is $P = \eta TP_{\text{exc}}(\Delta\Omega/4\pi) \approx (0.2)(0.4)(0.1)(0.02) \approx 1.6 \times 10^{-4}$. The experiment repetition rate is $R \approx 2 \times 10^3 \text{ s}^{-1}$, resulting in an entanglement generation rate $R_1 = PR \approx 0.3 \text{ s}^{-1}$.

Several improvements could significantly increase this yield. The repetition rate R could approach the excited-state spontaneous emission rate of $1/\tau_e \approx 10^8 \text{ s}^{-1}$. Using a fast, tailored laser pulse for the excitation could push P_{exc} towards unity, while eliminating multiple excitations. In addition, an imaging lens with a larger numerical aperture would improve the collection solid angle. A trade-off here is that higher collection efficiency comes at the cost of lower fidelity of the entangled state, which can be shown to vary as $F = 0.97 - 0.24(\Delta\Omega/4\pi)$ for $\Delta\Omega \ll 4\pi$. An alternative is to surround the ion with an optical cavity^{25,26} that would allow the collection of most of the photons scattered in each experiment and effectively make $\Delta\Omega/4\pi$ approach unity without sacrificing fidelity. It is important to note that this cavity need not be in the strong-coupling regime, as it affects only the entanglement efficiency, not the fidelity.

The atom-photon entanglement observed here may allow the generation of entangled states of remotely located trapped ion qubits, as follows. Consider two distant trapped ion qubits, each of which becomes entangled with their respective photons in the above fashion. When the emitted photons are mode-selected and combined on a beam splitter, appropriate coincidence measurements of the photons ensure the entanglement of the two trapped ion memories^{27–29}. The success probability for this twin event may be small (of order P^2), but when the entanglement eventually occurs after many trials, success is known, and this entanglement can be used for further scalable quantum information processing⁹. Given the low success probability P per trial in the current set-up, the rate of generating the two-ion entanglement would be $R_2 = P^2 R \approx 10^{-4} \text{ s}^{-1}$. However, with the improvements of fast excitation discussed above, and employing better optics, this rate could approach $R_2 \approx 10^3 \text{ s}^{-1}$ (assuming a 100-MHz experiment repetition rate). Finally, if reasonable-quality cavities were to be installed around each atomic qubit, this rate could approach $R_2 \approx 10^6 \text{ s}^{-1}$. $\square$

Methods

Polarization state of spontaneous emission along arbitrary axis

The polarization state of photons spontaneously emitted from an atomic dipole depends upon the change in angular momentum Δm of the atom along the dipole axis, and the direction of the emission, following the appropriate dipole radiation pattern. For $\Delta m = 0$, the (unnormalized) polarization state of a spontaneously emitted photon is $|I_0\rangle = -\sin\theta|\hat{\theta}\rangle$ and for $\Delta m = \pm 1$, the states are $|I_{\pm 1}\rangle = \frac{e^{\pm i\varphi}}{\sqrt{2}}(\cos\theta|\hat{\theta}\rangle \pm i|\hat{\varphi}\rangle)$, where θ and φ are the spherical polar and azimuthal angles of the emitted photon with respect to the dipole axis, and $\hat{\theta}$ and $\hat{\varphi}$ are their associated spherical coordinate unit vectors. Note that when a photon is emitted perpendicular to the dipole axis ($\theta = \pi/2$), the $\Delta m = \pm 1$ radiation is linearly polarized and orthogonal to the $\Delta m = 0$ radiation. For multiple decay channels of spontaneous emission to different states of atomic angular momentum $|S_{\Delta m_i}\rangle$, the resulting (unnormalized) state of photon and atom is $\Psi = \sum_{\Delta m_i} C_{\Delta m_i} |I_{\Delta m_i}\rangle |S_{\Delta m_i}\rangle$, where $C_{\Delta m_i}$ are atomic Clebsch-Gordon (CG) coefficients.

In the current experiment, spontaneous emission occurs via a $\Delta m = -1$ transition to atomic state $|S_{-1}\rangle = |1\rangle$ and via a $\Delta m = 0$ transition to atomic state $|S_0\rangle = |1\rangle$, with identical CG coefficients, resulting in the final (normalized) state of the emitted photon and atom:

$$\Psi = \frac{-\sqrt{2}\sin\theta|\hat{\theta}\rangle|\uparrow\rangle + e^{-i\varphi}\cos\theta|\hat{\theta}\rangle|\downarrow\rangle - ie^{-i\varphi}|\hat{\varphi}\rangle|\downarrow\rangle}{\sqrt{2+\sin^2\theta}}$$

Along a viewing axis orthogonal to the dipole ($\theta = \pi/2$), this state is highly entangled. In addition to this example, there are many other configurations that result in atom-photon entanglement.

Determining entanglement fidelity

The entanglement fidelity of an arbitrary two-qubit quantum state can be written as its

overlap with an appropriate maximally entangled two-qubit state³⁰. In the current experiment, the two qubits are represented by a 4×4 density matrix ρ with photon/atomic basis states $|H\uparrow\rangle$, $|H\downarrow\rangle$, $|V\uparrow\rangle$ and $|V\downarrow\rangle$. The entanglement fidelity with respect to the particular maximally entangled state $|\Psi_{\text{ME}}\rangle = \frac{1}{\sqrt{2}}(|H\uparrow\rangle + |V\downarrow\rangle)$ is given by:

$$F = \langle \Psi_{\text{ME}} | \rho | \Psi_{\text{ME}} \rangle = \frac{1}{2}(\rho_{H\uparrow, H\uparrow} + \rho_{V\downarrow, V\downarrow} + \rho_{H\uparrow, V\downarrow} + \rho_{V\downarrow, H\uparrow})$$

where $\rho_{IIS, IIS'} = \langle IIS | \rho | IIS' \rangle$ with $I = H$ or V and $S = \uparrow$ or $\downarrow$. (This expression also holds for any maximally entangled target state by appropriately redefining the basis states.) The first two terms in this expression are the measured correlation probabilities of detecting state $|H\rangle$ with $|\uparrow\rangle$ and state $|V\rangle$ with $|\downarrow\rangle$. The last two coherence terms can be determined by repeating the experiment while independently rotating each qubit through a polar angle of $\pi/2$ in the Bloch sphere before measurement. The rotated quantum state is then given by $\tilde{\rho} = R_{\pi/2}(\phi) \rho R_{\pi/2}^\dagger(\phi)$, where $R_{\pi/2}(\phi)$ is a $\pi/2$ polar rotation operator for both qubits with relative phase ϕ . We find that:

$$\begin{aligned} e^{-i\phi}\rho_{H\uparrow, V\downarrow} + e^{i\phi}\rho_{V\downarrow, H\uparrow} &= \tilde{\rho}_{H\uparrow, H\uparrow} - \tilde{\rho}_{V\downarrow, V\downarrow} - \tilde{\rho}_{H\downarrow, H\downarrow} - \tilde{\rho}_{V\uparrow, V\uparrow} - (e^{-i\phi}\rho_{V\downarrow, H\uparrow} + e^{i\phi}\rho_{H\uparrow, V\downarrow}) \\ &\geq \tilde{\rho}_{H\uparrow, H\uparrow} + \tilde{\rho}_{V\downarrow, V\downarrow} - \tilde{\rho}_{H\downarrow, H\downarrow} - \tilde{\rho}_{V\uparrow, V\uparrow} - 2\sqrt{\rho_{H\downarrow, H\downarrow}\rho_{V\uparrow, V\uparrow}} \end{aligned}$$

so that a lower bound on the entanglement fidelity can be expressed in terms of diagonal density matrix elements in the original and rotated basis (with ϕ set to zero):

$$F \geq \frac{1}{2}(\rho_{H\uparrow, H\uparrow} + \rho_{V\downarrow, V\downarrow} - 2\sqrt{\rho_{H\downarrow, H\downarrow}\rho_{V\uparrow, V\uparrow}} + \tilde{\rho}_{H\uparrow, H\uparrow} + \tilde{\rho}_{V\downarrow, V\downarrow} - \tilde{\rho}_{H\downarrow, H\downarrow} - \tilde{\rho}_{V\uparrow, V\uparrow})$$

These diagonals are expressed in terms of the measured probabilities as:

$$\rho_{IIS, IIS} = P(S|I)P(I) \text{ and } \tilde{\rho}_{IIS, IIS} = \tilde{P}(S|I)\tilde{P}(I)$$

For fidelities $F > 0.5$, the underlying quantum state is entangled²⁴.

Received 19 December 2003; accepted 30 January 2004; doi:10.1038/nature02377.

- DiVincenzo, D. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).
- Duan, L.-M., Lukin, M., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Bouwmeester, D., Ekert, A. & Zeilinger, A. (eds) *Quantum Cryptography, Quantum Teleportation, Quantum Computation* (Springer, Springer, 2000).
- Gheri, K., Ellinger, K., Pellizzari, T. & Zoller, P. Photon-wavepackets as flying quantum bits. *Fortschr. Phys.* **46**, 401–415 (1998).
- Haroche, S., Raimond, J. M. & Brune, M. in *Experimental Quantum Computation and Information* (eds de Martini, F. & Brune, M.) 3–36 (Proc. Int. School of Physics Enrico Fermi, course CXLVIII, IOS Press, Amsterdam, 2002).
- Kuhn, A. & Rempe, G. in *Experimental Quantum Computation and Information* (eds de Martini, F. & Monroe, C.) 37–66 (Proc. Int. School of Physics Enrico Fermi, course CXLVIII, IOS Press, Amsterdam, 2002).
- McKeever, J. et al. State-insensitive cooling and trapping of single atoms in an optical cavity. *Phys. Rev. Lett.* **90**, 133602 (2003).
- Briegleb, H.-J., Duer, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Duan, L.-M., Blinov, B. B., Moehring, D. L. & Monroe, C. Scalable trapped ion quantum computation with a probabilistic ion-photon mapping. Preprint at (<http://www.arxiv.org/quant-ph/0401020>) (2004).
- Freedman, S. J. & Clauser, J. F. Experimental test of local hidden variables theories. *Phys. Rev. Lett.* **28**, 938–941 (1972).
- Aspect, A., Grangier, P. & Roger, G. Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedanken experiment: a new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91–94 (1982).
- Eichmann, U. et al. Young's interference experiment with light scattered from two atoms. *Phys. Rev. Lett.* **70**, 2359–2362 (1993).
- DeVoe, R. G. & Brewer, R. G. Observation of superradiant and subradiant spontaneous emission of two trapped ions. *Phys. Rev. Lett.* **76**, 2049–2052 (1996).
- Kuzmich, A., Mandel, L. & Bigelow, N. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
- Julsgaard, B., Kozhekin, A. & Polzik, E. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
- Kuzmich, A. et al. Generation of nonclassical photon pairs for scalable quantum communication with atomic ensembles. *Nature* **423**, 731–734 (2003).
- van der Wal, C. H. et al. Atomic memory for correlated photon states. *Science* **301**, 196–200 (2003).
- Wineland, D. J. et al. Experimental issues in coherent quantum manipulation of trapped atomic ions. *NIST J. Res.* **103**, 259–328 (1998).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Sørensen, A. & Mølmer, K. Quantum computation with ions in thermal motion. *Phys. Rev. Lett.* **82**, 1971–1975 (1999).
- García-Ripoll, J. J., Zoller, P. & Cirac, J. I. Speed optimized two-qubit gates with laser coherent control techniques for ion trap quantum computing. *Phys. Rev. Lett.* **91**, 157901 (2003).
- Blatt, R. & Zoller, P. Quantum jumps in atomic systems. *Eur. J. Phys.* **9**, 250–256 (1988).
- Badziąg, P., Horodecki, M., Horodecki, P. & Horodecki, R. Local environment can enhance fidelity of quantum teleportation. *Phys. Rev. A* **62**, 012311 (2000).
- Sackett, C. A. et al. Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Guthorlein, G., Keller, M., Hayasaka, H., Lange, W. & Walther, H. A single ion as a nanoscopic probe of an optical field. *Nature* **414**, 49–51 (2001).
- Mundt, A. B. et al. Coupling a single atomic quantum bit to a high finesse optical cavity. *Phys. Rev. Lett.* **89**, 103001 (2002).
- Duan, L.-M. & Kimble, J. Efficient engineering of multiatom entanglement through single-photon detections. *Phys. Rev. Lett.* **90**, 253601 (2003).

28. Simon, C. & Irvine, W. Robust long-distance entanglement and a loophole-free Bell test with ions and photons. *Phys. Rev. Lett.* **91**, 110405 (2003).
29. Cabrillo, C., Cirac, J. I., Garcia-Fernandez, P. & Zoller, P. Creation of entangled states of distant atoms by interference. *Phys. Rev. A* **59**, 1025–1033 (1999).
30. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).

Acknowledgements We acknowledge discussions with M. Madsen, P. Haljan, M. Acton and D. Wineland, and thank R. Miller for assistance in building the trap apparatus. This work was supported by the National Security Agency, the Advanced Research and Development Activity, under Army Research Office contract, and the National Science Foundation Information Technology Research Division.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.B. (bblinov@umich.edu).

Supramolecular dendritic liquid quasicrystals

Xiangbing Zeng¹, Goran Ungar¹, Yongsong Liu¹, Virgil Percec²,
Andrés E. Dulcey² & Jamie K. Hobbs³

¹Department of Engineering Materials, University of Sheffield, Sheffield S1 3JD, UK

²Roy & Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, USA

³H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK

A large number of synthetic and natural compounds self-organize into bulk phases exhibiting periodicities on the 10^{-8} – 10^{-6} metre scale¹ as a consequence of their molecular shape, degree of amphiphilic character and, often, the presence of additional non-covalent interactions. Such phases are found in lyotropic systems² (for example, lipid–water, soap–water), in a range of block copolymers³ and in thermotropic (solvent-free) liquid crystals⁴. The resulting periodicity can be one-dimensional (lamellar phases), two-dimensional (columnar phases) or three dimensional ('micellar' or 'bicontinuous' phases). All such two- and three-dimensional structures identified to date obey the rules of crystallography and their symmetry can be described, respectively, by one of the 17 plane groups or 230 space groups. The 'micellar' phases have crystallographic counterparts in transition-metal alloys, where just one metal atom is equivalent to a 10^3 – 10^4 -atom micelle. However, some metal alloys are known to defy the rules of crystallography and form so-called quasicrystals, which have rotational symmetry other than the allowed two-, three-, four- or six-fold symmetry⁵. Here we show that such quasiperiodic structures can also exist in the scaled-up micellar phases, representing a new mode of organization in soft matter.

Research on bulk nanoscale self-assembly of organic matter is partly motivated by the fact that such complex structures may serve as scaffolds for photonic materials⁶ and other nanoarrays, or as precursors for mesoporous ceramics or elements for molecular electronics. Larger biological objects, such as cylinder-like or sphere-like viruses, also pack on similar macrolattices⁷.

Dendrons and dendrimers (tree-like molecules⁸) are proving particularly versatile in generating periodic nanostructures (Fig. 1). Two micellar lattices, with space groups $Im\bar{3}m$ (body-centred cubic, b.c.c.)⁹, and $Pm\bar{3}n$ (refs 10, 11), have been established. An analogue of the $Im\bar{3}m$ phase has also been observed in block copolymers¹², and that of the $Pm\bar{3}n$ phase in lyotropic liquid crystals¹³. Recently, a complex three-dimensional (3D) tetragonal

lattice (space group $P4_2/mnm$) was found, having 30 self-assembled micelles in the unit cell (Fig. 1f)¹⁴.

In many dendron systems, thermal transitions between the phases in Fig. 1 occur. The master sequence $Col_h \rightarrow Pm\bar{3}n \rightarrow P4_2/mnm \rightarrow Im\bar{3}m$ is obeyed with increasing temperature; in only a handful of cases are all these phases displayed in the same material. In a number of compounds, however, an additional unidentified phase has been observed below any other 3D phase but above Col_h . A small-angle X-ray powder diffractogram of this phase, recorded on dendron **I** (Fig. 1g), is shown in Fig. 2a. The synthesis of **I** is described in ref. 15 and Supplementary Information, where this compound is labelled $[3,4,5-(3,5)^2]12G_3CH_2OH$. Other compounds that show the X-ray signature of this phase include $(4-3,4,5-3,5)12G_2CH_2OH$, $[4-(3,4,5)^2]12G_2COOH$, $[3,4-(3,5)^2]12G_3COOH$, $[3,4-(3,5)^2]12G_3CH_2OH$, $[3,4-(3,4,5)^2]12G_3CH_2OH$ (ref. 15), polyoxazolines with tapered side groups containing alkyl chains of different lengths¹⁶, as well as certain salts of 3,4,5-*tris*-(*n*-alkoxy)benzoic acid¹⁷.

On heating, compound **I** shows the following phase sequence: room temperature $\rightarrow X \rightarrow 71^\circ C \rightarrow P4_2/mnm \rightarrow 72^\circ C \rightarrow$ isotropic liquid, while on cooling phase X forms directly from the liquid (Supplementary Information). This allowed us to grow mono-domains of the unknown phase. That phase X is a quasicrystal is revealed by the distinctive but crystallographically forbidden 12-fold symmetry of the small-angle X-ray single-crystal pattern (Fig. 2b). When the sample is rotated around the 12-fold axis with the incident beam perpendicular to the axis, the diffraction pattern repeats every 30° . One such pattern is shown in Fig. 2c, where the Ewald sphere cuts through a pair of strong reflections in Fig. 2b. The structure of this liquid quasicrystal (LQC) is periodic in the direction of the 12-fold axis, but quasiperiodic in the plane perpendicular to it.

In contrast to normal 3D periodic structures, five instead of three basis vectors are needed for indexing the diffraction peaks of a dodecagonal quasicrystal¹⁸. Four of the vectors, q^1 , q^2 , q^3 and q^4 ,

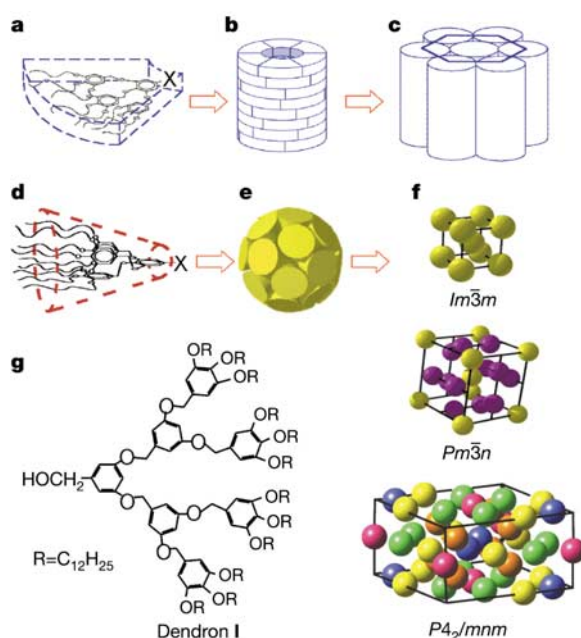


Figure 1 Self-assembly of wedge-shaped molecules. **a**, Dendrons with fewer tethered chains adopt a flat slice-like shape (X is a weakly binding group). **b**, The slices stack up and form cylindrical columns, which assemble on a two-dimensional hexagonal columnar (Col_h) lattice (**c**). **d**, Dendrons with more end-chains assume a conical shape. **e**, The cones assemble into spheres, which pack on three different 3D lattices (**f**) with symmetries $Im\bar{3}m$, $Pm\bar{3}n$ and $P4_2/mnm$. **g**, Structure of compound **I** studied in this work.

are perpendicular to the 12-fold axis (Fig. 2b). q^5 , along the 12-fold axis, is indicated in Fig. 2c. Thus each diffraction peak q is indexed by five integers ($n^1 n^2 n^3 n^4$ and n^5) where $q = \sum_{i=1}^5 n^i q^i$. Tentative indices of some of the diffraction peaks are given in Fig. 2.

The structure of the LQC is closely related to those of the $Pm\bar{3}n$ and $P4_2/mnm$ phases. Where these phases form from the LQC on

heating, their {002} reflections appear at the same position as the {00002} reflections of the LQC. Thus the periodicity of the LQC along the 12-fold axis is the same as that of $Pm\bar{3}n$ along any of the three cubic axes, and $P4_2/mnm$ along c . This suggests that, like the above phases, the LQC is also micellar. With increasing temperature, the observed d -spacings of the LQC decrease proportionally. This isotropic shrinkage suggests a phase containing isometric objects.

$Pm\bar{3}n$, $P4_2/mnm$ and $Im\bar{3}m$ (b.c.c.) phases that are observed in dendrimers all have their structural equivalents in transition metals or alloys (for example, $Pm\bar{3}n$: Cr_3Si ; $P4_2/mnm$: $Fe_{46}Cr_{54}$ and β -uranium; b.c.c.: α -iron). The dendrimer 'atoms' (micelles) have volumes several thousand times larger than real atoms, and so do their unit cells. Both $Pm\bar{3}n$ and $P4_2/mnm$ phases belong to the family of tetrahedrally close packed (t.c.p.) structures of spherical objects, or Frank-Kasper phases¹⁹. In a t.c.p. structure, any four neighbouring spheres pack tetrahedrally, which is locally the densest packing. However, regular tetrahedral interstices are incompatible with long-range order. Such frustration leads to the complexity of

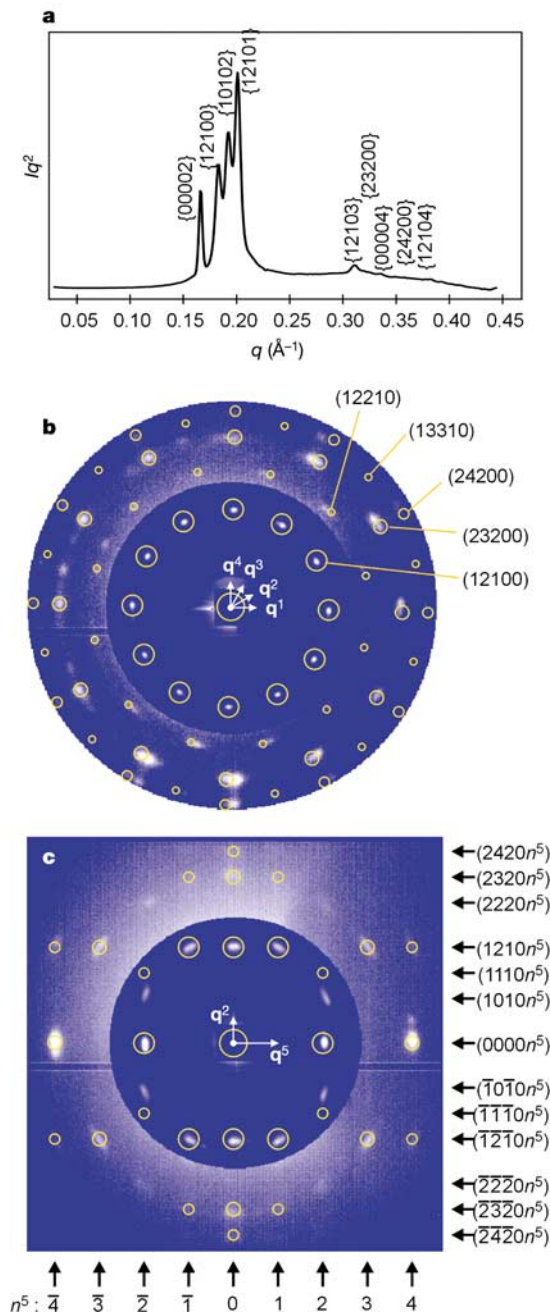


Figure 2 Experimental and simulated X-ray diffraction patterns of the LQC. **a**, Powder diffraction pattern of compound **I** recorded at 70 °C. **b**, Precession single-crystal pattern along the 12-fold axis. The intensities of outer diffraction peaks in **b** and **c** are scaled up by 100. **c**, Single-crystal diffraction pattern perpendicular to the 12-fold axis, cutting through one of the six pairs of strong diffraction spots in **b**. The pattern repeats itself every 30° when the sample is rotated around the 12-fold axis. All single-crystal patterns were recorded at room temperature. Simulated diffraction patterns are superimposed, with the reflections represented by circles whose area is proportional to log(amplitude). The apparent deviation in position of diffraction spots in the outer region of the diffraction pattern in **c** is due to the existence of other domains in the sample (see Supplementary Information).

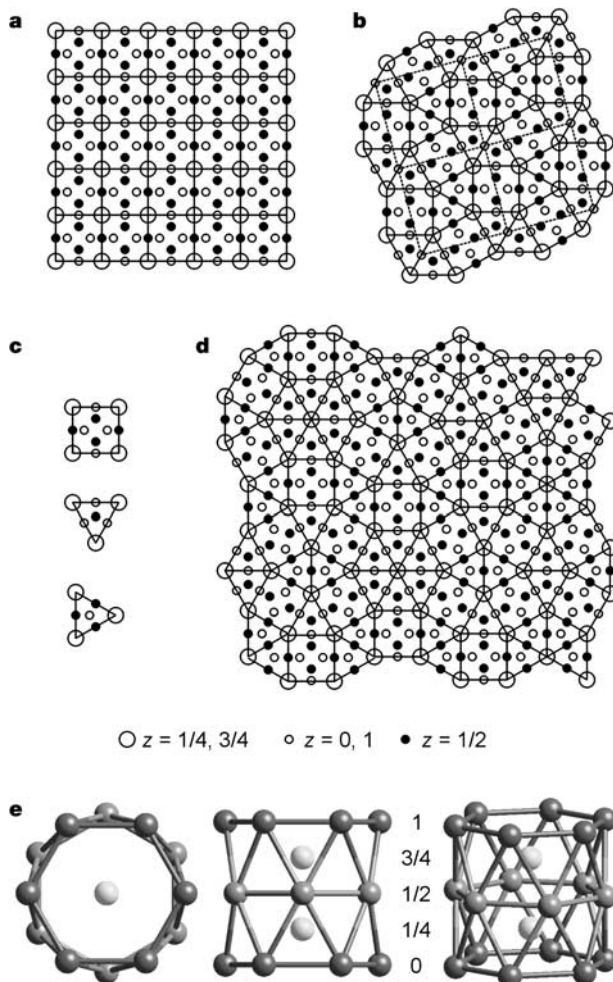


Figure 3 Packing of spheres in the LQC and other related t.c.p. structures. All can be generated from 2D tilings consisting of only squares and equilateral triangles ('sparse' nets, elevations $z = 1/4$ and $3/4$, large open circles). These tiles are 'decorated' with spheres at $z = 0$ (small open circles) and $1/2$ (filled circles), forming the 'dense' nets. **a**, Sphere packing in $Pm\bar{3}n$. **b**, Sphere packing in $P4_2/mnm$. **c**, Only three different decorated tiles, one square and two triangular, are used to generate both $Pm\bar{3}n$ and $P4_2/mnm$ structures. **d**, The same tiles, when arranged quasiperiodically, will generate the model of the LQC with 12-fold symmetry. The lattice constant, that is, the length of the tile edge, as well as the periodicity along the 12-fold axis, is 81.4 Å at room temperature. **e**, Two ideal hexagonal antiprisms stacked along the $\bar{1}2$ symmetry axis. There is a distorted hexagonal antiprism at each node of the square-triangular tiling in **a**, **b** and **d**.

t.c.p. structures. Ultimately, it leads to quasicrystals.

Dodecagonal quasicrystals have been found so far in five transition-metal systems^{20–23}. In all but one case, dodecagonal structures were studied by electron microscopy or electron diffraction. The interpretation of electron diffraction is complicated by multiple scattering. In the one case where quasicrystals sufficiently large for single-crystal X-ray diffraction were obtained, the stacking along the 12-fold axis was different from that in other dodecagonal phases, including the present LQC²³. As X-ray data can be compared with models more unequivocally than electron diffraction data, we proceed to construct a model of the LQC. We follow the idea that dodecagonal structures can be generated by appropriate ‘decoration’ of a quasiperiodic square-triangular tiling^{18,21,23}.

$Pm\bar{3}n$ and $P4_2/mnm$ structures are characterized by alternating densely and sparsely populated layers, as shown in Fig. 3a and b, respectively. The nets generated by connecting the nearest neighbours in the sparsely populated layers (large circles) are equivalent to tilings that cover an infinite plane using only squares and equilateral triangles. The full 3D crystal structure is then generated by ‘decoration’, that is, by associating with each tile a column, periodic along c , containing individual micelles. Only three kinds of tiles, one square and two triangular (Fig. 3c), are needed to construct the $Pm\bar{3}n$ and $P4_2/mnm$ structures. (In the actual $P4_2/mnm$ phase, the position of the micelles is somewhat different from those in an idealized structure made up of decorated tiles (Fig. 3c). Similarly, it is expected that in the real quasicrystal, the micelles would also be somewhat away from the ideal positions assumed by our starting model (Fig. 3d)).

Using quasiperiodic tiling, the same elements can be used to construct models of the LQC (Fig. 3d). For ways of generating such tilings, see refs 24 and 25. Such models readily explain the 12-fold symmetry, and the equal periodicities in the third dimension for the LQC, $Pm\bar{3}n$ and $P4_2/mnm$ phases. Simulated diffraction patterns based on the quasiperiodic square-triangular tiling of ref. 25 are superimposed on the experimental ones in Fig. 2b and c (for details of the simulation, see Supplementary Information). Even the unrefined starting model gives a reasonable explanation of the positions and relative intensities of experimental diffraction spots (Fig. 2b and c). Nevertheless, discrepancies in intensities remain (for example, {10102} and {22202} reflections are stronger than expected). Improvement of the model is progressing, as the reliability of acquired diffraction data allows it.

Atomic force microscope images of the LQC phase on glass show that the structure is periodic in one direction but not in the other, in agreement with a quasicrystalline structure where the 12-fold axis lies parallel to the glass surface. Moreover, they support the ‘micellar’ nature and the layered structure of the LQC (Supplementary Information).

To explain why spherical aggregates of dendrons pack on different 3D lattices, mathematical functions defining molecular shapes ideally suited to typical lattices have been calculated¹⁴. In a different approach²⁶, micelles were approximated by hard cores and soft alkyl coronas. Assuming that thermal transitions are primarily driven by the laterally expanding alkyl chains, the first micellar phase to which a columnar dendrimer would transform on heating is expected to be the one with the minimal coronal surface area. Such a structure ought to coincide with the solution of the Kelvin problem of the ultimate equilibrium dry foam (minimum surface area per bubble). The b.c.c. structure proposed by Kelvin²⁷ was challenged only recently by a calculation showing that the $Pm\bar{3}n$ structure provides a better solution²⁸. The fact that, of all known micellar phases in self-assembled dendrimers, the $Pm\bar{3}n$ was always found at the lowest temperature was taken as experimental vindication of the new minimal surface solution²⁶. However, now that the updated phase sequence is $Col_h \rightarrow LQC \rightarrow Pm\bar{3}n \rightarrow P4_2/mnm \rightarrow$ b.c.c., the possibility is raised of dodecagonal quasicrystals providing a still better solution of the Kelvin problem.

Outside metal-based systems, non-crystallographic rotational symmetry has been observed only in thin films of smectic C twist grain boundary (TGBC) liquid crystals confined between glass plates with rubbed surfaces^{29,30}. q -fold symmetry is created by helical packing of smectic C grains, and lock-in transitions between different q -values were observed with changing temperature. In contrast, the present LQC is a bulk phase with intrinsic high-level quasiperiodic order inherent in the molecular architecture.

The tiling of planes normal to the 12-fold axis in the LQC (Fig. 3d) can be understood intuitively in the following way. There are six tetrahedrons packed around the vertical line connecting the two light-grey spheres at $z = 1/4$ and $3/4$ in Fig. 3e, taking two adjacent spheres at $z = 1/2$ as the other two tetrahedral apices. However, the number of perfect tetrahedrons that can fit around a common edge is 5.1. In different t.c.p. structures, including LQC, the inability to cover a flat surface with regular pentagons is resolved in different tilings of distorted hexagons. The problem is reversed in the case of curved surfaces; these cannot be tiled exclusively by hexagons. Thus in ‘buckyballs’ and footballs, a certain proportion of pentagons is required and, as the curvature radius decreases relative to tile size, this proportion increases until the all-pentagonal icosahedron is reached. Interesting examples are found in hexagonal and pentagonal packing modes of proteins in capsids (coats) of cylindrical and spherical viruses³¹.

Finally, the unusual symmetry of quasicrystals has been exploited in recent years for fabrication of photonic bandgap arrays. Owing to their high symmetry, two-dimensional (2D) quasicrystalline lattices are able to induce and widen the photonic bandgap³², preventing light within a range of wavelengths from propagating in any direction. More recently it was shown that light can be slowed down in a one-dimensional photonic quasicrystal that follows the Fibonacci sequence. To produce photonic ‘quasicrystals’ with photonic bandgaps in the visible light region, the distance of two neighbouring objects must be on the scale of several hundred nanometres. Whereas the results presented here show how the characteristic length of a self-assembled quasicrystal can be scaled up from a few ångströms in metal alloys to nearly 10 nm in supramolecular dendrimers, it may be possible to achieve a further increase of two orders of magnitude by using appropriately designed self-assembling dendrons, block copolymers or other soft sphere systems. □

Received 23 October 2003; accepted 26 January 2004; doi:10.1038/nature02368.

1. Supramolecular chemistry and self-assembly. *Science* **295** (special issue), 2395–2421 (2002).
2. Seddon, J. M. Lyotropic phase behaviour of biological amphiphiles. *Ber. Bunsenges. Phys. Chem.* **100**, 380–393 (1996).
3. Thomas, E. L., Anderson, D. M., Henke, C. S. & Hoffman, D. Periodic area-minimizing surfaces in block copolymers. *Nature* **334**, 598–601 (1988).
4. Tschierske, C. Micro-segregation, molecular shape and molecular topology partners for the design of liquid crystalline materials with complex mesophase morphologies. *J. Mater. Chem.* **11**, 2647–2671 (2001).
5. Janot, C. *Quasicrystals: A Primer* (Oxford Univ. Press, Oxford, 1992).
6. Lopes, W. A. & Jaeger, H. M. Hierarchical self-assembly of metal nanostructures on diblock copolymer scaffolds. *Nature* **414**, 735–738 (2001).
7. Horne, R. W. in *Molecular Plant Virology* Vol. 1 (ed. Davies, J. W.) 1–41 (CRC, Boca Raton, 1985).
8. Newkome, G. R., Moorefield, C. N. & Vogtle, F. *Dendrimers and Dendrons* (Wiley-VCH, Weinheim, 2001).
9. Yeardley, D. J. P., Ungar, G., Percec, V., Holerca, M. N. & Johansson, G. Spherical supramolecular minidendrimers self-organized in an “inverse micellar”-like thermotropic body-centered cubic liquid crystalline phase. *J. Am. Chem. Soc.* **122**, 1684–1689 (2000).
10. Balagurusamy, V. S. K., Ungar, G., Percec, V. & Johansson, G. Rational design of the first spherical supramolecular dendrimers self-organized in a novel thermotropic cubic liquid-crystalline phase and the determination of their shape by X-ray analysis. *J. Am. Chem. Soc.* **119**, 1539–1555 (1997).
11. Hudson, S. D. *et al.* Direct visualization of individual cylindrical and spherical supramolecular dendrimers. *Science* **278**, 449–452 (1997).
12. Bates, F. S., Cohen, R. E. & Berney, C. V. Small-angle neutron scattering determination of macro-lattice structure in a polystyrene polybutadiene diblock co-polymer. *Macromolecules* **15**, 589–592 (1982).
13. Luzzati, V., Vargas, R., Mariani, P., Gulik, A. & Delacroix, H. Cubic phases of lipid-containing systems—Elements of a theory and biological connotations. *J. Mol. Biol.* **229**, 540–551 (1993).
14. Ungar, G., Liu, Y. S., Zeng, X. B., Percec, V. & Cho, W.-D. Giant supramolecular liquid crystal lattice. *Science* **299**, 1208–1211 (2003).
15. Percec, V., Cho, W.-D., Ungar, G. & Yeardley, D. J. P. Synthesis and structural analysis of two

- constitutional isomeric libraries of AB₂-based monodendrons and supramolecular dendrimers. *J. Am. Chem. Soc.* **123**, 1302–1315 (2001).
16. Percec, V., Holerca, M. N., Uchida, S., Yearley, D. J. P. & Ungar, G. Poly(oxazoline)s with tapered minidendritic side groups as models for the design of synthetic macromolecules with tertiary structure. *Biomacromolecules* **2**, 729–740 (2001).
 17. Percec, V. *et al.* Exploring and expanding the three-dimensional structural diversity of supramolecular dendrimers with the aid of libraries of alkali metals of their AB₃ minidendritic carboxylates. *Chem. Eur. J.* **8**, 1106–1117 (2002).
 18. Gähler, F. in *Quasicrystalline Materials* (eds Janot, C. & Dubois, J. M.) 272–284 (World Scientific, Singapore, 1988).
 19. Frank, F. C. & Kasper, J. S. Complex alloy structures regarded as sphere packing. I. Definitions and basic principles. *Acta Crystallogr. B* **11**, 184–190 (1958).
 20. Ishimasa, T., Nissen, H.-U. & Fukano, Y. New ordered state between crystalline and amorphous in Ni-Cr particles. *Phys. Rev. Lett.* **55**, 511–513 (1985).
 21. Chen, H., Li, D. X. & Kuo, K. H. New type of two-dimensional quasicrystal with twelvefold rotational symmetry. *Phys. Rev. Lett.* **60**, 1645–1648 (1988).
 22. Yoshida, K., Yamada, T. & Taniguchi, Y. Long-period tetragonal lattice formation by solid-state alloying at the interfaces of Bi-Mn double-layer thin-films. *Acta Crystallogr. B* **45**, 40–45 (1989).
 23. Conrad, M., Krumeich, F. & Harbrecht, B. A dodecagonal quasicrystalline chalcogenide. *Angew. Chem. Int. Edn Engl.* **37**, 1383–1386 (1998).
 24. Stampfli, P. A dodecagonal quasiperiodic lattice in two dimensions. *Helv. Phys. Acta* **59**, 1260–1263 (1986).
 25. Baake, M., Klitzing, R. & Schlottmann, M. Fractally shaped acceptance domains of quasiperiodic square-triangle tilings with dodecagonal symmetry. *Physica A* **191**, 554–558 (1992).
 26. Zihler, P. & Kamien, R. D. Maximizing entropy by minimizing area: Towards a new principle of self-organization. *J. Phys. Chem. B* **105**, 10147–10158 (2001).
 27. Thomson, W. On the division of space with minimum partitional area. *Phil. Mag.* **24**, 503–514 (1887).
 28. Weaire, D. (ed.) *The Kelvin Problem: Foam Structures of Minimal Surface Area* (Taylor & Francis, London, 1997).
 29. Navailles, L., Barois, P. & Nguyen, H. T. X-ray measurement of the twisted grain boundary angle in the liquid crystal analog of the Abrikosov phase. *Phys. Rev. Lett.* **71**, 545–548 (1993).
 30. Navailles, L., Pindak, R., Barois, P. & Nguyen, H. T. Structural study of the smectic-C twisted grain boundary phase. *Phys. Rev. Lett.* **74**, 5224–5227 (1995).
 31. Bruinsma, R. F., Gelbart, W. M., Reguera, D., Rudnick, J. & Zandi, R. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* **24**, 248101 (2003).
 32. Zoorob, M. E., Charlton, M. D. B., Parker, G. J., Baumberg, J. J. & Netti, M. C. Complete photonic bandgaps in 12-fold symmetric quasicrystals. *Nature* **404**, 740–743 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Gleeson and P. Baker for assistance with X-ray diffraction experiments. We are grateful to P. A. Heiney, T. C. Lubensky and R. D. Kamien for reading the draft manuscript and for their suggestions. We acknowledge CCLRC for providing synchrotron beamtime. The synthesis part of the work was supported by the NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.U. (g.ungar@sheffield.ac.uk).

Links between salinity variation in the Caribbean and North Atlantic thermohaline circulation

Matthew W. Schmidt¹, Howard J. Spero¹ & David W. Lea²

¹Department of Geology, University of California, Davis, California 95616, USA

²Department of Geological Sciences and Marine Science Institute, University of California, Santa Barbara, California 93106, USA

Variations in the strength of the North Atlantic Ocean thermohaline circulation have been linked to rapid climate changes¹ during the last glacial cycle through oscillations in North Atlantic Deep Water formation and northward oceanic heat flux^{2–4}. The strength of the thermohaline circulation depends on the supply of warm, salty water to the North Atlantic, which, after losing heat to the atmosphere, produces the dense water masses that sink to great depths and circulate back south². Here we analyse two Caribbean Sea sediment cores, combining Mg/Ca palaeothermometry with measurements of oxygen isotopes in

foraminiferal calcite in order to reconstruct tropical Atlantic surface salinity^{5,6} during the last glacial cycle. We find that Caribbean salinity oscillated between saltier conditions during the cold oxygen isotope stages 2, 4 and 6, and lower salinities during the warm stages 3 and 5, covarying with the strength of North Atlantic Deep Water formation⁷. At the initiation of the Bølling/Allerød warm interval, Caribbean surface salinity decreased abruptly, suggesting that the advection of salty tropical waters into the North Atlantic amplified thermohaline circulation and contributed to high-latitude warming.

Today, most of the North Atlantic's subtropical gyre water circulates through the Caribbean Sea before it is transported to the subpolar regions of the North Atlantic via the Gulf Stream⁸. Net evaporation exceeds precipitation in the Atlantic, resulting in freshwater removal of $\sim 0.35 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ from the Atlantic basin⁹. Because of their influence on North Atlantic surface salinity, the tropical and subtropical Atlantic play an important part in regulating North Atlantic Deep Water (NADW) formation. However, unlike the Gulf of Mexico and North Atlantic, Caribbean salinity is not significantly affected by freshwater runoff and therefore surface salinity primarily reflects the evaporation/precipitation ratio over the western tropical Atlantic. Hence, changes in tropical

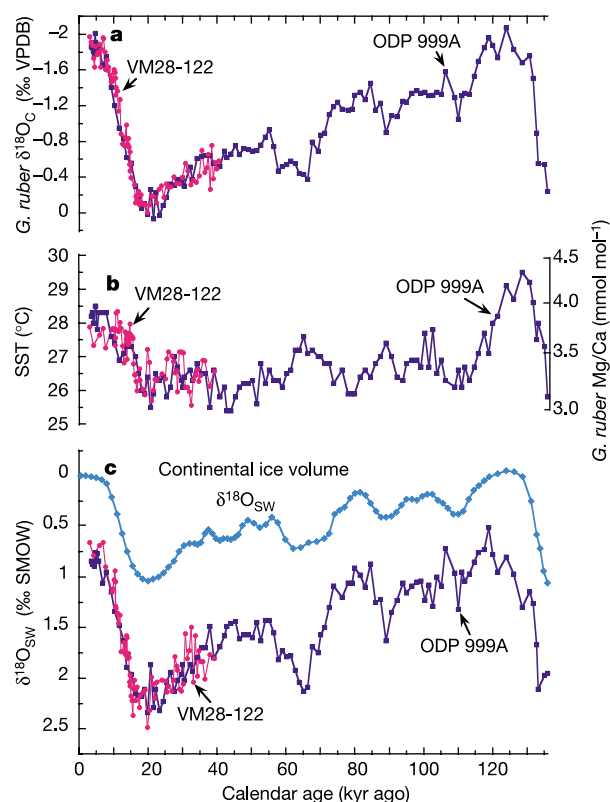


Figure 1 Temperature and $\delta^{18}\text{O}_{\text{SW}}$ variation in the western Caribbean Sea during the past 136 kyr. **a**, Colombian basin $\delta^{18}\text{O}_{\text{C}}$ and **b**, Mg/Ca–SST records from ODP 999A ($12^\circ 45' \text{ N}$, $78^\circ 44' \text{ W}$; 2,827 m; 4 cm kyr^{-1} sedimentation rate) and VM28-122 ($11^\circ 34' \text{ N}$, $78^\circ 25' \text{ W}$; 3,623 m; 4 cm kyr^{-1} sedimentation rate during the Holocene and LGM, $10\text{--}15 \text{ cm kyr}^{-1}$ sedimentation rate during the deglaciation), based on the planktic foraminifer *G. ruber* (white). **c**, Computed $\delta^{18}\text{O}_{\text{SW}}$ calculated from the Mg/Ca-derived SST and $\delta^{18}\text{O}_{\text{C}}$ using $T(\text{in } ^\circ\text{C}) = 16.5 - 4.80(\delta_{\text{C}} - (\delta_{\text{W}} - 0.27))$ (ref. 14). The continental ice-volume $\delta^{18}\text{O}_{\text{SW}}$ reconstruction¹⁵ is shown for comparison. Note that the amplitude of the calculated $\delta^{18}\text{O}_{\text{SW}}$ change in the Colombian basin is considerably greater than the global $\delta^{18}\text{O}_{\text{SW}}$ change due to ice volume alone.

atmospheric circulation have a profound effect on Caribbean surface salinity, which could in turn affect the salinity and density structure of the entire North Atlantic.

The $^{18}\text{O}/^{16}\text{O}$ ratio of sea water covaries linearly with surface salinity, making it one of the most direct proxies for estimating salinity in the modern ocean¹⁰. On glacial timescales, the $\delta^{18}\text{O}$ value for sea water ($\delta^{18}\text{O}_{\text{SW}}$) is also affected by variations in continental ice volume because the formation of glacial ice preferentially removes H_2^{16}O from the ocean. The $\delta^{18}\text{O}$ of foraminiferal calcite ($\delta^{18}\text{O}_{\text{C}}$) is controlled by temperature and $\delta^{18}\text{O}_{\text{SW}}$, so $\delta^{18}\text{O}_{\text{SW}}$ can be computed if temperature is determined independently. Initial attempts to compute North Atlantic $\delta^{18}\text{O}_{\text{SW}}$ used indirect sea surface temperature (SST) estimates based on faunal transfer functions⁴. Given the uncertainties in determining the precise magnitude and timing of SST change from transfer functions and the fact that the computed temperature did not necessarily reflect conditions at the time the foraminiferal shell locked in its $\delta^{18}\text{O}_{\text{C}}$ value, the resulting surface salinity reconstructions are only qualitative. Here we combine Mg/Ca measurements (a proxy for the temperature of calcification) and $\delta^{18}\text{O}$ analyses of shells from the surface-dwelling foraminifera *Globigerinoides ruber* (white) from the western Caribbean Colombian basin at ODP Site 999A and VM28-122 (see Supplementary Information for core location map) to produce the first continuous record of Caribbean $\delta^{18}\text{O}_{\text{SW}}$ during the last 136 kyr (Fig. 1). The most significant advantage of Mg palaeothermometry over other temperature determinations is that Mg/Ca is measured on the same organism and mineral phase that carries the $\delta^{18}\text{O}$ information. Hence, there is no spatial or temporal ambiguity in the incorporation of the two controlling parameters^{5,6} and ambient $\delta^{18}\text{O}_{\text{SW}}$ can be computed directly.

Age models for ODP 999A and VM28-122 were developed using eight new and six previously published accelerator mass spectroscopy

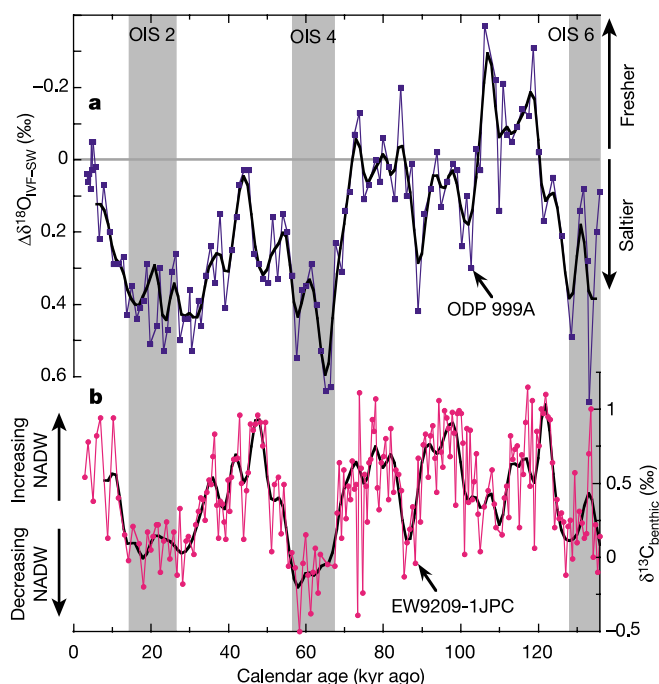


Figure 2 A comparison of local $\delta^{18}\text{O}_{\text{SW}}$ change in Colombian basin surface water and estimated variation in NADW formation over the past 136 kyr. **a**, $\Delta\delta^{18}\text{O}_{\text{IVF-SW}}$ (salinity excess) for ODP 999A represents regional hydrologic changes in the Caribbean through the last glacial cycle. The bold line is a low-pass (5 kyr) filter with a 5-point filter weight running through the raw data (fine line). **b**, Benthic $\delta^{13}\text{C}$ record (filtered with a low-pass (5 kyr) filter with a 10-point filter weight) from Ceara Rise core EW9209-1JPC (5°N , 43°W ; 4,056 m) indicates times of reduced NADW formation (lower $\delta^{13}\text{C}$)⁷.

(AMS) ^{14}C dates from *G. ruber* and correlation of the *G. ruber* $\delta^{18}\text{O}_{\text{C}}$ records to SPECMAP¹¹ and to estimates of sea level change from coral U–Th dates¹² (see Supplementary Information).

We convert measured Mg/Ca ratios to SST using a depth-corrected *G. ruber* Mg/Ca–SST calibration for the tropical Atlantic¹³ (Fig. 1b). To compute $\delta^{18}\text{O}_{\text{SW}}$, temperature is then removed from the $\delta^{18}\text{O}_{\text{C}}$ record using a temperature– $\delta^{18}\text{O}$ relationship that has been field-calibrated for use with *G. ruber* (white)¹⁴ (Fig. 1c). The component of $\delta^{18}\text{O}_{\text{SW}}$ that is due to changes in regional hydrology during the last glacial cycle was then calculated by subtracting the influence of continental ice volume¹⁵ and subtracting the modern Caribbean surface $\delta^{18}\text{O}_{\text{SW}}$ value, producing the ice-volume-free residual, $\Delta\delta^{18}\text{O}_{\text{IVF-SW}}$, presented as the deviation from modern conditions (Fig. 2a).

The *G. ruber* $\delta^{18}\text{O}_{\text{C}}$ record from ODP 999A displays a glacial–interglacial amplitude of 2‰ (Fig. 1a). The Holocene and the interglacial Oxygen Isotope Stage (OIS) 5e attain a minimum value of -2.0‰ and the Last Glacial Maximum (LGM) peaks at 0‰ . The $\delta^{18}\text{O}_{\text{C}}$ record in VM28-122 is nearly identical to ODP 999A through the past 40 kyr. VM28-122 shows virtually all of the detail contained in a $10\text{--}20\text{ cm kyr}^{-1}$ *G. ruber* $\delta^{18}\text{O}_{\text{C}}$ record from core M35003-4 in the Tobago basin¹⁶, just east of the Caribbean Sea (Fig. 3a).

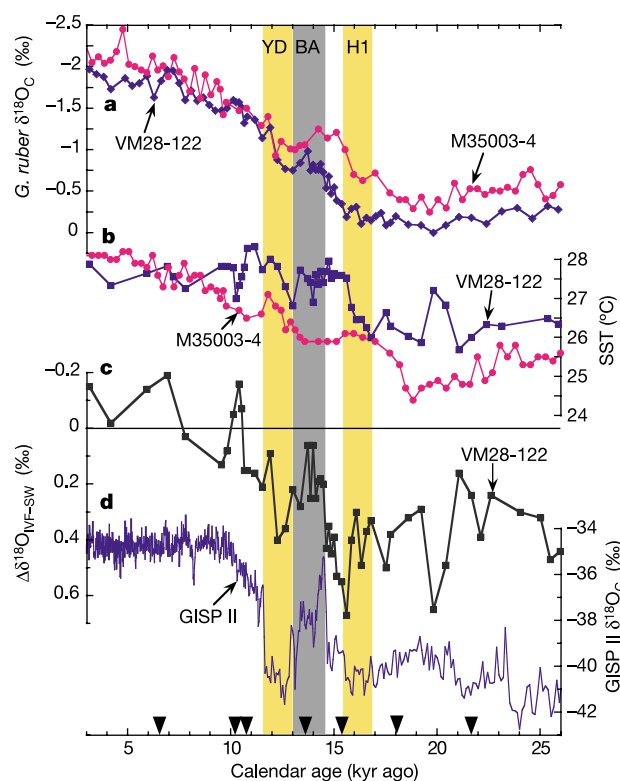


Figure 3 Temperature and salinity variation in the Colombian Basin and in the western tropical Atlantic over the last deglacial. Records of **a**, $\delta^{18}\text{O}_{\text{C}}$ and **b**, SST from VM28-122 and from Tobago basin core M35003-4 (12°N , 61°W ; 1,299 m) (ref. 16) are compared. $\delta^{18}\text{O}_{\text{C}}$ is from *G. ruber* (white) in VM28-122 and from *G. ruber* (pink) in M35003-4. The SST record for VM28-122 is derived from Mg/Ca, whereas the M35003-4 SST record is alkenone-based. The differences in the $\delta^{18}\text{O}_{\text{C}}$ and SST records from VM28-122 and M35003-4 may be due to the influence of the low-salinity pool associated with the Amazon and Orinoco river outflow on the Tobago basin or to a southward shift in the glacial NEC. **c**, The computed $\Delta\delta^{18}\text{O}_{\text{IVF-SW}}$ record from VM28-122 represents regional surface salinity change in the Caribbean. Note that the rapid salinity decrease in the Caribbean (14.6 kyr ago) corresponds to an increase in the GISP II $\delta^{18}\text{O}$ record (**d**)¹⁹ at the beginning of the Bolling/Allerød warm interval. Arrows on the bottom axis indicate VM28-122 intervals with AMS ^{14}C dates. Heinrich Event 1 (H1), the Bolling/Allerød (BA) and the Younger Dryas (YD) are indicated.

The Mg/Ca records from ODP 999A and VM28-122 yield a late Holocene SST of $\sim 28.0^\circ\text{C}$ and show that the western Caribbean was $\sim 2.5^\circ\text{C}$ cooler during the LGM (Fig. 1b). These late Holocene SST values agree with the average modern SST for the Colombian basin¹⁷ and the glacial–interglacial temperature difference is consistent with previous estimates for the Caribbean¹⁸. The SST record from ODP 999A also shows that OIS 4 was $\sim 1.2^\circ\text{C}$ warmer than OIS 2 and 3. The VM28-122 Mg/Ca record, which has a sedimentation rate of over 15 cm kyr^{-1} during the mid-deglaciation, shows a pause in SST rise associated with the Bølling/Allerød interval¹⁹ (Fig. 3b). This pattern of SST change is similar to that observed from alkenone data in core M35003-4 (ref. 16 and Fig. 3b), but differs from a coastal Cariaco basin SST reconstruction based on Mg/Ca palaeothermometry that indicates significant cooling during the Younger Dryas²⁰. Both the Tobago and Cariaco basins yield lower LGM SSTs than the western Caribbean. Because the LGM was a period of enhanced carbonate preservation in the Caribbean, it is unlikely that these LGM temperature differences are due to the influence of dissolution on Mg/Ca in our data. Furthermore, our reported Mg/Ca ratios show a statistically significant negative correlation with average shell weight, suggesting that partial dissolution has not affected our Mg/Ca–SST reconstruction. Rather, the SST difference between the western Caribbean and the eastern sites may be due to a southward shift in the glacial North Equatorial Current (NEC)²¹.

Subtracting the influence of SST on $G. ruber$ $\delta^{18}\text{O}_\text{C}$ in ODP 999A and VM28-122, we obtain the salinity proxy, $\delta^{18}\text{O}_\text{SW}$ (Fig. 1c). The late Holocene yields an average $\delta^{18}\text{O}_\text{SW}$ of 0.8‰ , in close agreement with modern annual $\delta^{18}\text{O}_\text{SW}$ values for the western Caribbean²². The $\delta^{18}\text{O}_\text{SW}$ record for the past 136 kyr has a glacial–interglacial amplitude of $\sim 1.5\text{‰}$, exceeding estimates of $\delta^{18}\text{O}_\text{SW}$ change due to ice volume by 0.5‰ (ref. 15). Removing the influence of ice volume¹⁵ and subtracting the modern Caribbean surface $\delta^{18}\text{O}_\text{SW}$ value of 0.8‰ yields $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ (Fig. 2a). Positive values indicate increased surface salinity due to the excess removal of H_2O during evaporation. In ODP 999A, $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ was more positive than modern seawater by ~ 0.5 to 0.6‰ during cold glacial intervals OIS 2, 4 and 6. In contrast, warm OIS 3 and 5 values of $\delta^{18}\text{O}_\text{SW}$ are indistinguishable from the average modern $\delta^{18}\text{O}_\text{SW}$ value in this region (for example, $\Delta\delta^{18}\text{O}_\text{IVF-SW} \approx 0$). If we assume the freshwater $\delta^{18}\text{O}$ value in the tropical Atlantic did not change significantly during the LGM²³, then the modern western Caribbean $\delta^{18}\text{O}_\text{SW}$ to surface salinity relationship, $\delta^{18}\text{O}_\text{SW} = 0.22 \times (\text{surface salinity}) - 6.95$ (ref. 22), is valid for the last glacial cycle. Our glacial $\delta^{18}\text{O}_\text{SW}$ values therefore indicate that Caribbean salinities were 2.3 to 2.7 p.s.u. higher than modern salinities, after the influence of ice volume on oceanic salinity has been subtracted. Previous research also indicates increased western tropical Atlantic surface salinity during glacial periods²⁴, suggesting a regional increase in glacial surface salinity that extended beyond the Caribbean.

Climate in the modern Colombian basin is characterized by two distinct seasons: a warm, wet season in late summer when the Intertropical Convergence Zone (ITCZ) is located farthest to the north, and a cool, dry season during boreal winter when the ITCZ migrates south²⁵. As a result, surface salinity varies seasonally by ~ 0.7 p.s.u. (ref. 17). The increased glacial salinities calculated in ODP 999A (Fig. 2a) are therefore consistent with modelling²⁶ and sedimentological data²⁷ suggesting a southward shift in the ITCZ during periods of reduced NADW formation.

It is generally accepted that cold periods in the North Atlantic are associated with reduced NADW formation³. On orbital timescales, benthic foraminiferal $\delta^{13}\text{C}$ oscillations from the western tropical Atlantic reflect the relative strength of NADW (high $\delta^{13}\text{C}$) and Antarctic Bottom Water (low $\delta^{13}\text{C}$) production. Compared with the $\delta^{13}\text{C}$ record from western tropical Atlantic Ocean core EW9209-1 JPC⁷ over the last glacial cycle (Fig. 2b), reconstructed $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ from ODP 999A demonstrates that elevated Caribbean salinities occurred when NADW production was reduced. The combined

evidence of arid atmospheric conditions²⁷ and covariation between benthic $\delta^{13}\text{C}$ and $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ during the last glacial cycle suggests that increased glacial Caribbean salinities were a function of both tropical Atlantic hydrological changes and a general decrease in flow rate through the Caribbean that resulted in a weaker Gulf Stream²⁸.

Comparison of the VM28-122 $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ record with the GISP II $\delta^{18}\text{O}$ record¹⁹ (Fig. 3c, d) reveals that Caribbean surface salinity decreased rapidly at the onset of the Bølling/Allerød interval. The $\delta^{18}\text{O}_\text{SW}$ transition from glacial to modern salinity (as determined by the midpoint of the VM28-122 $\Delta\delta^{18}\text{O}_\text{IVF-SW}$ transition) occurs during a period of increased sedimentation rate and is coincident with the onset of the Bølling/Allerød at 14.6 kyr ago. During the brief period of Bølling/Allerød warmth in the North Atlantic, Colombian basin salinity reduced significantly, reaching a minimum between 14.5 and 13.7 kyr ago. Both the initiation and termination of this surface-salinity minimum are bracketed by AMS dates. Sedimentological evidence from the Cariaco basin also suggests a rapid northward migration of the ITCZ during the Bølling/Allerød, resulting in enhanced rainfall in the Caribbean region²⁷. This salinity decrease was preceded by a brief period of very salty conditions in the Caribbean that could be related to the collapse of thermohaline circulation during Heinrich Event 1 (Fig. 3c). The observed pattern of millennial-scale surface salinity change suggests that cold stadial events are characterized by a drier western tropical Atlantic combined with reduced surface flow through the Caribbean. These surface-salinity reconstructions agree with Cariaco basin sedimentological data but contradict an earlier interpretation that interstadials are periods of increased salinity in the western tropical Atlantic²⁷.

Recent modelling studies suggest the initiation of the Bølling/Allerød warming and concurrent strengthening of North Atlantic thermohaline circulation was triggered by either a reduction in Southern Ocean salinity due to meltwater pulse 1A²⁹ or warming and sea-ice retreat³⁰. It has been difficult to explain how the warm Bølling/Allerød interval could result from a reinvigoration of thermohaline circulation when North Atlantic surface warming, sea-ice retreat and continental deglaciation should have produced a negative density anomaly that would have reduced deep convection. We hypothesize that excess salt advected from the Caribbean at ~ 14.6 kyr ago (Fig. 3c) may have helped to offset the low-salinity conditions in the North Atlantic. Model results suggest that the Indian Ocean was an additional source of salt to the North Atlantic at the start of the deglaciation³⁰. When combined, these two sources of salt would serve as a density amplifier for deep thermohaline circulation, thereby offsetting the effect of rising temperature and melting ice.

It is clear that the tropical oceans play a key role in glacial–interglacial climate transitions. Because the accumulation of excess salt in the tropical North Atlantic ultimately impacts the density of high-latitude surface waters, it affects North Atlantic climate through its influence on thermohaline circulation. On the basis of our results, surface waters in the Caribbean became exceptionally salty during glacial intervals over the last 136 kyr. Amplification of North Atlantic thermohaline circulation associated with warm phases during glacial cycles may be critically dependent on the delivery of salty waters that accumulate in the Caribbean during cold periods. These results suggest that the tropical hydrologic cycle has a direct role in regulating rapid climate change. □

Methods

Sediment from each core interval was disaggregated in ultrapure water, sieved and dried at 45°C . To minimize intraspecific variation in shell geochemistry at each core interval, specimens of *G. ruber* s.s. (white) were only collected from the 250–350 μm size fraction. On the basis of preliminary analyses of the intraspecific oxygen isotope variability among *G. ruber* within core intervals, we used 25 shells per stable isotope analysis for all downcore measurements. Samples were collected at 5 cm resolution (ODP 999A) and 2 cm resolution (VM28-122). Samples for stable isotopes were sonicated in methanol for 5–10 s, roasted under vacuum at 375°C for 30 min, and analysed on a Fisons Optima IRMS, using

an Isocarb common acid bath autocarbonate system at 90 °C at the University of California Davis.

Mg/Ca ratios were measured on foraminifera collected from the same population and size fraction that was used for the stable isotope analyses. Approximately 600 µg of material per sample (~55 shells) was cleaned for trace and minor element analysis⁵. Briefly, samples underwent a multi-step process consisting of initial rinses in ultrapure water, followed by treatments with hot reducing and oxidizing solutions, transfers into new acid-leached micro-centrifuge vials, and finally leaches with a dilute ultrapure acid solution. All sample cleaning was conducted in laminar flow benches under trace-metal-clean conditions. Samples were then dissolved and analysed on a Finnigan Element-2 ICPMS at the University of California Santa Barbara using established procedures⁵. A full suite of trace- and minor-element measurements were made on each sample including Ca, Mg, Sr, Na, Cd, Ba, La, Ce, Nd, Eu, Lu, U, Al, Mn and Fe.

Elemental ratios of Fe/Ca and Al/Ca were used to monitor cleaning efficacy. Analyses with anomalously high Fe/Ca or Al/Ca ratios and/or with recovery of less than 20% after cleaning were rejected. Although Fe/Ca ratios gradually increased with depth in both cores, Al/Ca ratios remained uniformly low, typically <7 µmol mol⁻¹ in ODP 999A and <14 µmol mol⁻¹ in VM28-122, indicating that Mg contamination associated with detrital sediment in cleaned samples was not an issue at either site. Although Fe/Ca and Mn/Ca were higher in VM28-122, Mg/Ca did not correlate with Fe/Ca, Al/Ca or Mn/Ca in either core, indicating that the mineral phases containing these metals in the cleaned samples did not affect Mg/Ca ratios in the foraminiferal calcite.

Error analysis

Analytical precision for the δ¹⁸O_C measurements is better than ±0.06‰. The pooled standard deviation of replicate Mg/Ca analyses from ODP 999A was ±1.7% (1 s.d., d.f. = 110). The pooled standard deviation of replicate Mg/Ca analyses from VM28-122 was ±2.4% (1 s.d., d.f. = 129). The overall precision of replicates in this study is slightly better than other tropical cores (typically ~3%)^{5,20}, most probably reflecting the stability of the water column in the Colombian basin through the last glacial cycle. Standard deviation for the δ¹⁸O_{SW} residual was calculated to be ±0.2‰, using Monte Carlo methodology that assumed a 1σ normal distribution in the δ¹⁸O_C and Mg/Ca measurements and in the Mg/Ca–SST and δ¹⁸O–SST calibrations.

Received 24 October 2003; accepted 19 January 2004; doi:10.1038/nature02346.

- Boyle, E. A. Is ocean thermohaline circulation linked to abrupt stadial/interstadial transitions? *Quat. Sci. Rev.* **19**, 255–272 (2000).
- Broecker, W. S., Bond, G., Klas, M., Bonani, G. & Wolffli, W. A salt oscillator in the glacial Atlantic? The concept. *Paleoceanography* **5**, 469–478 (1990).
- Keigwin, L. D. & Boyle, E. A. Surface and deep ocean variability in the northern Sargasso Sea during marine isotope stage 3. *Paleoceanography* **14**, 164–170 (1999).
- Duplessy, J. C. *et al.* Changes in surface salinity of the North Atlantic Ocean during the last deglaciation. *Nature* **358**, 485–488 (1992).
- Lea, D. W., Pak, D. K. & Spero, H. J. Climate impact of Late Quaternary equatorial Pacific sea surface temperature variations. *Science* **289**, 1719–1724 (2000).
- Elderfield, H. & Ganssen, G. Past temperature and δ¹⁸O of surface ocean waters inferred from foraminiferal Mg/Ca ratios. *Nature* **405**, 442–445 (2000).
- Curry, W. B. & Oppo, D. W. Synchronous high-frequency oscillations in tropical sea surface temperatures and North Atlantic deep water production during the last glacial cycle. *Paleoceanography* **12**, 1–14 (1997).
- Johns, W. E., Townsend, T. L., Frantoni, D. M. & Wilson, W. D. On the Atlantic inflow to the Caribbean Sea. *Deep-Sea Res.* **1** **49**, 211–243 (2002).
- Broecker, W. S., Peng, T. H., Jouzel, J. & Russell, G. The magnitude of global fresh-water transports of importance to ocean circulation. *Clim. Dyn.* **4**, 73–79 (1990).
- Craig, H. & Gordon, L. I. In *Stable Isotopes in Oceanographic Studies and Paleotemperatures* (ed. Tongiorgi, E.) 9–130 (CNR, Pisa, 1965).
- Bassinot, F. C. *et al.* The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
- Chappell, J. *et al.* Reconciliation of late Quaternary sea levels derived from coral terraces at Huon Peninsula with deep sea oxygen isotope records. *Earth Planet. Sci. Lett.* **141**, 227–236 (1996).
- Dekens, P. S., Lea, D. W., Pak, D. K. & Spero, H. J. Core top calibration of Mg/Ca in tropical foraminifera: refining paleotemperature estimation. *Geochem. Geophys. Geosyst.* **3**, no. 1022 (2002).
- Thunell, R., Tappa, E., Pride, C. & Kincaid, E. Sea-surface temperature anomalies associated with the 1997–1998 El Niño recorded in the oxygen isotope composition of planktonic foraminifera. *Geology* **27**, 843–846 (1999).
- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
- Ruhlemann, C., Mulitza, S., Muller, P. J., Wefer, G. & Zahn, R. Warming of the tropical Atlantic ocean and slowdown of thermohaline circulation during the last deglaciation. *Nature* **402**, 511–514 (1999).
- Conkright, M. *et al.* NODC Internal Report at (<http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NODC/WOA98/>) (NODC, Silver Springs, Maryland, 1998).
- Hastings, D. W., Russell, A. D. & Emerson, S. R. Foraminiferal magnesium in *Globigerinoides sacculifer* as a paleotemperature proxy. *Paleoceanography* **13**, 161–169 (1998).
- Grootes, P. M. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10³ to 10⁵-year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).
- Lea, D. W., Pak, D. K., Peterson, L. C. & Hughes, K. A. Synchronicity of tropical high latitude Atlantic temperatures over the last glacial termination. *Science* **301**, 1361–1364 (2003).
- Vink, A. *et al.* Shifts in the position of the North Equatorial Current and rapid productivity changes in the western Tropical Atlantic during the last glacial. *Paleoceanography* **16**, 1–12 (2001).
- Watanabe, T. A., Winter, A. & Oba, T. Seasonal changes in sea surface temperature and salinity during the Little Ice Age in the Caribbean Sea deduced from Mg/Ca and 18O/16O ratios in corals. *Mar. Geol.* **173**, 21–35 (2001).
- Jouzel, J., Hoffmann, G., Koster, R. D. & Masson, V. Water isotopes in precipitation: data/model comparison for present-day and past climates. *Quat. Sci. Rev.* **19**, 363–379 (2000).

- Dürkoop, A., Hale, W., Mulitza, S., Patzold, J. & Wefer, G. Late Quaternary variations of sea surface salinity and temperature in the western tropical Atlantic: Evidence from δ¹⁸O of *Globigerinoides sacculifer*. *Paleoceanography* **12**, 764–772 (1997).
- Stidd, C. K. The use of eigenvectors for climate estimates. *J. Appl. Meteor.* **6**, 255–264 (1967).
- Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54**, 251–267 (2002).
- Peterson, L. C., Haug, G. H., Hughes, K. A. & Rohl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last Glacial. *Science* **290**, 1947–1951 (2000).
- Lynch-Stieglitz, J., Curry, W. B. & Slowey, N. Weaker Gulf stream in the Florida Straits during the Last Glacial Maximum. *Nature* **402**, 644–648 (1999).
- Weaver, A. J., Saaenko, O. A., Clark, P. U. & Mitrovica, J. X. Meltwater pulse 1A from Antarctica as a trigger of the Bolling/Allerod warm interval. *Science* **299**, 1709–1713 (2003).
- Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424**, 532–536 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Ocean Drilling Program (ODP) and the LDEO Deep-Sea Sample Repository for core samples. Laboratory assistance from D. Pak and mass spectrometer operation by G. Paradis and D. Winter were critical to the success of this study. We also thank A. Droxler for providing a suite of his ODP 999A samples, and J. Kennett and A. Russell for their comments and suggestions. Funding for this research was provided by a USSSP Schlanger Ocean Drilling Fellowship to M.W.S. and the US National Science Foundation (H.J.S. and D.W.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W.S. (schmidt@geology.ucdavis.edu).

Seismic reflection imaging of two megathrust shear zones in the northern Cascadia subduction zone

Andrew J. Calvert

Department of Earth Sciences, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada

At convergent continental margins, the relative motion between the subducting oceanic plate and the overriding continent is usually accommodated by movement along a single, thin interface known as a megathrust¹. Great thrust earthquakes occur on the shallow part of this interface where the two plates are locked together². Earthquakes of lower magnitude occur within the underlying oceanic plate, and have been linked to geochemical dehydration reactions caused by the plate's descent^{3–7}. Here I present deep seismic reflection data from the northern Cascadia subduction zone that show that the inter-plate boundary is up to 16 km thick and comprises two megathrust shear zones that bound a >5-km-thick, ~110-km-wide region of imbricated crustal rocks. Earthquakes within the subducting plate occur predominantly in two geographic bands where the dip of the plate is inferred to increase as it is forced around the edges of the imbricated inter-plate boundary zone. This implies that seismicity in the subducting slab is controlled primarily by deformation in the upper part of the plate. Slip on the shallower megathrust shear zone, which may occur by aseismic slow slip, will transport crustal rocks into the upper mantle above the subducting oceanic plate and may, in part, provide an explanation for the unusually low seismic wave speeds that are observed there^{8,9}.

The Cascadia subduction zone, where the oceanic Juan de Fuca plate descends beneath the overlying North American plate, extends 1,100 km from northern California to northern Vancouver Island. Sedimentary rocks originally deposited on the oceanic plate are

scraped off and accreted to the western edge of the continent; the initiation of this process is indicated by the deformation front in Fig. 1. The intersection of the deformation front with the southwest-trending Nootka transform fault, which is characterized by high levels of seismicity, marks the approximate northern limit of the subduction zone. Beneath the western and eastern edges of Vancouver Island, earthquakes within the oceanic plate occur in two well-defined, ~50-km-wide, geographic bands that merge further south. These earthquakes have been attributed to dehydration embrittlement of the oceanic mantle⁷ and the gradual transformation of basalt to eclogite^{6,7}; this latter process may begin at depths as shallow as 40 km (ref. 6).

West of Vancouver Island, seismic refraction and normal incidence reflection surveys show that the top of the subducting oceanic crust lies immediately beneath a seismic reflector that dips landward at between 12 and 25 km depth^{10,11} ('JdF' in Figs 2a and 3). Beneath southern Vancouver Island and its margins, seismic reflection surveys also identify two regionally extensive structures: a thick zone of reflectors, denoted by the letter E, and a deeper, probably individual, reflector, identified by the letter F (refs 12–15). On a composite reflection section across the northern Cascadia forearc (Fig. 2a), the E zone appears as a sequence of anastomosing reflections that extends over 1.0 to 3.2 s (corresponding to a thickness of 3–10 km), dips at 4–15°, and reaches a depth of at least 45 km east of Vancouver Island. The second reflection, F, is observed intermittently, between ~0.5 and 2.0 s (~2–6 km) later than the deepest reflector of the E sequence. The short duration of this reflection implies that it originates either from a thin, <2-km-thick, region, or at a single interface.

In offshore surveys, the F reflection can be traced seaward and upward towards the reflection from the top of the subducting plate west of Vancouver Island. As the F reflection shallows it approaches to within 0.5 s (<2 km) of the deepest of the E reflections, defining a wedge-like geometry (Fig. 2b). If the top of the subducting plate were located at the base of the E reflectors, then the most plausible interpretation of the F reflector would be as the oceanic Moho¹⁶, implying that the descending oceanic crust is <2 km thick in places. However, this conclusion is inconsistent with observations in the deep ocean basin where the Moho of the incoming oceanic plate is observed 1.8–2.2 s after the top of the igneous crust¹⁷. In fact, the Moho of the subducting plate is located by wide-angle reflections⁷ recorded in the SHIPS^{18,19} survey and occurs ~6 km deeper than the F reflector (Fig. 2b), consistent with modelling of gravity data²⁰ and analysis of teleseismic P to S wave conversions recorded near central Vancouver Island²¹. I therefore interpret the F reflector to be the top of the subducting plate.

The E reflections dip shallowly to the east-northeast (ENE) at approximately 4° across southern Vancouver Island. As the reflections reach the eastern edge of the island, their dip increases sharply to ~15° and they continue into the uppermost mantle with this dip. In contrast, the F reflection from the top of the subducting plate dips to the ENE at ~15° near the western edge of Vancouver Island, and then flattens out further to the east. It is not possible to identify unequivocally the shorter F reflection segments near the east coast of Vancouver Island owing to elevated levels of coherent noise; however, it seems unlikely that the F reflection would crosscut the E reflections. Therefore the F reflector must also increase in dip as it presumably merges into the deepest of the E reflectors near the eastern edge of Vancouver Island. Thus identified, the E and F reflectors define a ~110-km-wide duplex structure: the F reflection, which marks the top of the subducting plate, is the floor thrust with a ramp-flat-ramp geometry, and the E reflectors form a vertically distributed roof thrust (Fig. 4). In the region bounded by the E and F reflectors, P wave velocities (derived by three-dimensional (3D) tomographic inversion of first arrivals from local earthquakes and first arrivals of the SHIPS survey⁹) vary laterally between 6,800 and 7,200 m s⁻¹ (Fig. 2b). These velocity values are consistent with

crustal rocks of both the oceanic plate and the lower continental plate.

Earthquake hypocentres relocated as part of the 3D tomographic inversion⁹ can be correlated accurately with the seismic reflection data and a corresponding vertical section extracted from the 3D velocity model. Earthquakes that occurred within 15 km of the seismic profiles have been projected laterally onto the profile and are shown in Fig. 2b, in which the reflection section has been converted to depth. The depth conversion used the extracted velocity model extended seaward with a combination of refraction velocities¹⁰ and stacking velocities from the reflection profile²², and removed the effects on deep reflectors of lateral velocity variations in the near-surface. Earthquakes can be divided into two main groups, upper and lower, depending on whether they are above or below the E reflectors, which appear to be notably aseismic¹⁶.

Earthquakes occur above the E reflectors to depths as great as 35 km, but the seismicity terminates abruptly at the shallowest of the E reflectors. This distribution of earthquakes suggests that the region of the lower continental crust in which the E reflectors are located either cannot sustain stress accumulation or is completely locked; the latter is unlikely given an estimated temperature at the E reflectors of ~400 °C (ref. 23). Landward of the locked part of the inter-plate boundary, which extends landward 60 km from the deformation front², accumulating stress is released through episodic, aseismic slow-slip events²⁴. Simultaneous non-volcanic tremors²⁵ have been approximately located at multiple depths over at least 15 km above the subducting plate²⁶. The existence of tremors at depths corresponding to the E reflectors, and the absence of conventional seismicity within the E reflectors, imply that slow slip is occurring here. Tremors occur at depths corresponding to the F reflector, suggesting that slip is also occurring at the top of the subducting plate. With the roof thrust of the duplex active, the

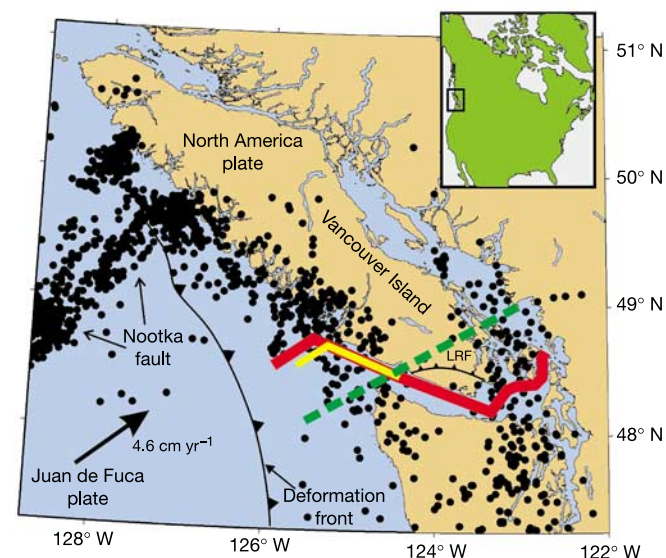


Figure 1 Map of the northern Cascadia subduction zone showing the distribution of inslab earthquakes (filled black circles) along the margin. The approximate northern limit of the Cascadia subduction zone is marked by the intersection of the deformation front with the Nootka fault, which is characterized by high levels of seismicity caused by relative motion between the Juan de Fuca plate and a small oceanic plate to the north. Earthquakes within the subducting slab occur predominantly along the western and eastern edges of Vancouver Island. The seismic lines combined to form the cross-section across the forearc region in Fig. 2 are shown in red, the azimuth of the simulated cross-section in green (dashed). The lines combined to form the section in Fig. 3 are shown in yellow. LRF, Leech River fault.

enclosed imbricated rocks will be transported downward (Fig. 4), perhaps into the forearc upper mantle, and this process may provide an explanation for the P wave velocities as low as $6,800 \text{ m s}^{-1}$ that have been observed there^{8,9}.

Near the west coast of Vancouver Island, numerous earthquakes, with a variety of focal mechanisms²⁷, occur beneath the top of the subducting oceanic crust where it dips most steeply; only one earthquake occurs above the top of the plate, being located in the imbrication enclosed within the duplex structure (Figs 2b, 3). As the top of the plate flattens, the number of earthquakes within the slab

decreases markedly even though dehydration of the crust should be occurring here⁶. As indicated in Fig. 1, there is another belt of inslab earthquakes near the east coast of Vancouver Island, and Fig. 2b shows that this belt corresponds to the location where the inferred dip of the subducting plate increases once more. I propose that the two belts of inslab seismicity (Fig. 1) are due either to flexure or to fragmentation of the oceanic plate as it subducts obliquely around the edges of the imbricated inter-plate boundary zone (Fig. 4). Dehydration reactions may reduce the effective normal stresses across faults⁷ and facilitate rupture when the subducting plate

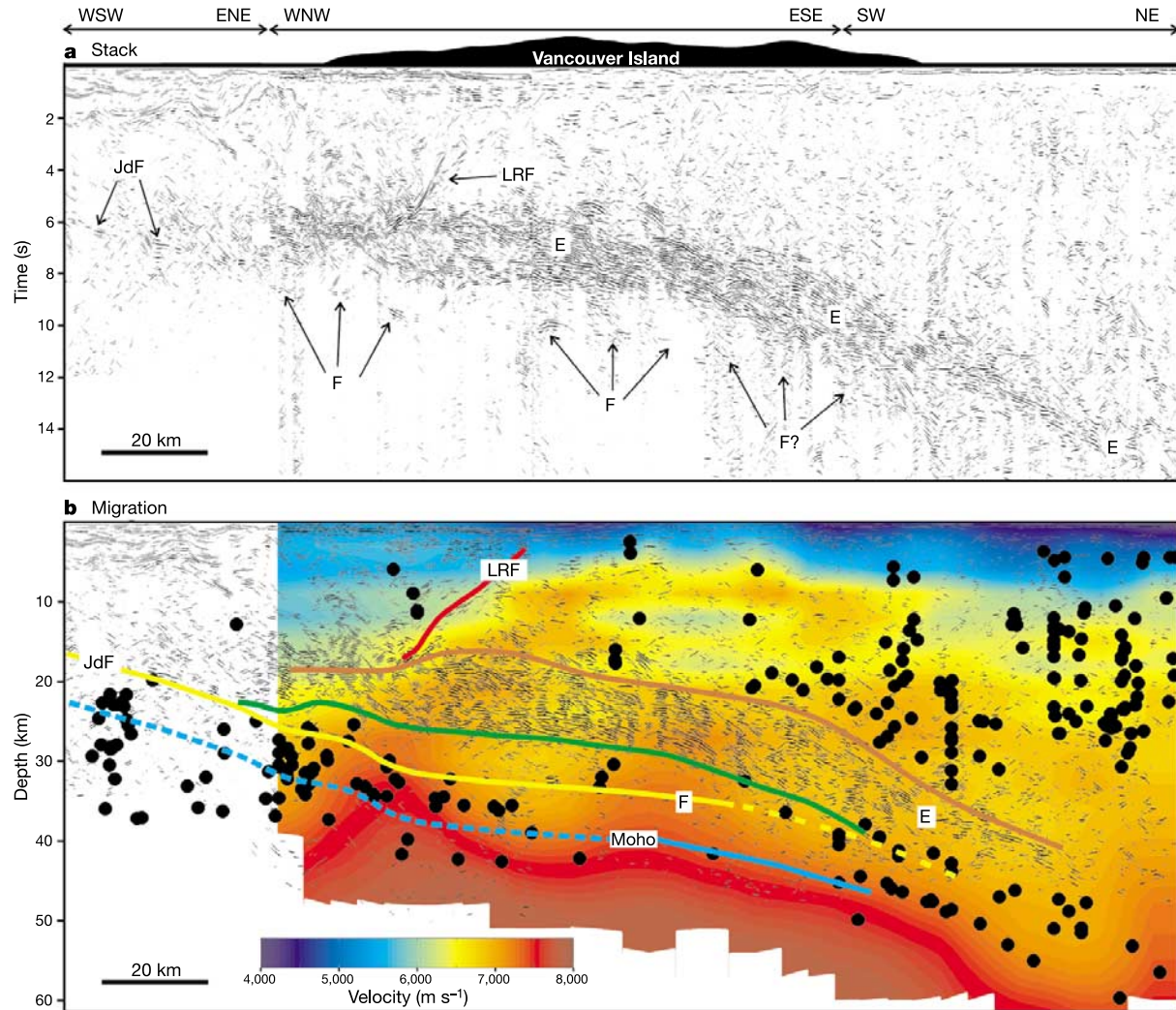


Figure 2 Composite seismic cross-section across the Cascadia forearc region in the vicinity of southern Vancouver Island. **a**, Unmigrated section showing the reflection, labelled JdF, from the top of the subducting Juan de Fuca plate west of Vancouver Island, the E reflection zone, and the F reflection which marks the top of the subducting Juan de Fuca plate further east. Reflections from the Leech River crustal fault, LRF, are truncated by the E reflections. The lack of continuity of the F reflector beneath the western half of Vancouver Island is largely due to the gain used for this display; on higher-gain displays, the F reflector is continuous over ~80% of the profile with breaks attributable to localized variations in near-surface basement topography. The cross-section was constructed by projecting individual seismic profiles onto an azimuth of 063° , which is a representative dip direction for the subducting plate near the seismic profiles. The apparent dips of reflections such as JdF, E and F, which dip shallowly in directions close to the projection azimuth, are close to true dip on the simulated section. The orientation of the original line segments used to construct the composite section are shown at the top. **b**, Migrated section superimposed on a display of P wave velocities derived by 3D tomographic

inversion of first arrivals from local earthquakes and wide-angle airgun shots around Vancouver Island⁹. The velocities were extracted from the 3D velocity model along the composite seismic profile shown by the red line in Fig. 1 and projected in the same way as the seismic reflection data. Earthquakes, which were relocated in the tomographic inversion, are shown by filled black circles. The near-linear alignment of east-dipping inslab earthquakes just above the oceanic Moho at 45–55 km depth may indicate delamination of the oceanic crust. Steeply dipping seismic reflectors oblique to the line of projection may be mispositioned; however, for the shallowly ~ENE-dipping reflectors interpreted here, any lateral positioning errors will probably be of the order of the accuracy of the hypocentre locations and the cell size used in the tomographic inversion, namely 3 km, or less. Oceanic Moho constrained by wide-angle reflections, blue; oceanic Moho assuming subducting crust is 6 km thick, dashed blue; top of subducting plate, yellow; deepest E reflection, green; shallowest E reflection, brown; Leech River fault, red. Vertical exaggeration is 1.5:1.

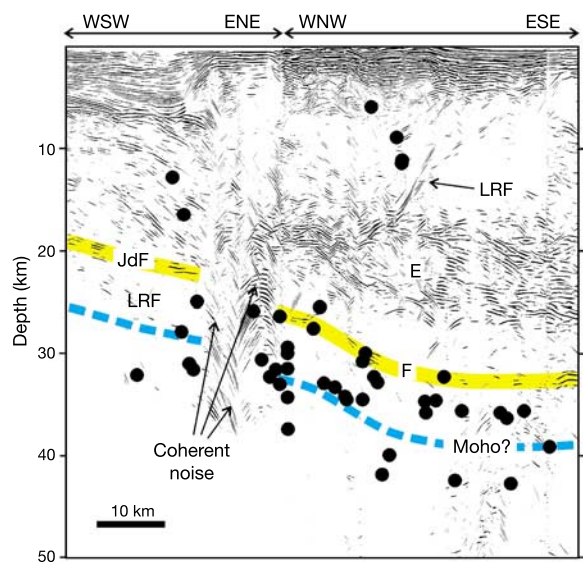


Figure 3 Composite seismic section near the west coast of Vancouver Island. The shallowest of the inslab earthquakes follow the geometry of the top of the subducting Juan de Fuca plate (yellow line). Inslab earthquakes occur predominantly where the dip of the subducting plate increases to 15°, and infrequently elsewhere. Earthquake hypocentres within 15 km of the seismic profiles are shown by black filled circles. The seismic data are unmigrated, but the position of the top of the plate is derived from a migrated section, indicating that the position of the interface does not move a great distance after migration owing to its relatively low dip. The orientation of the original line segments used to construct the composite section are shown at the top. Oceanic Moho assuming subducting crust is 6 km thick, dashed blue line. Vertical exaggeration is 1.5:1.

flexes, for example, near the west coast of Vancouver Island. At the eastern edge of Vancouver Island, the oceanic crust reaches 40 km depth where the transformation of basalt to eclogite begins⁶, and focal mechanisms are dominated by downdip tension²⁷. The consequent increase in density of the crust may facilitate the sudden increase in dip of the subducting slab here, and perhaps result in delamination of the crust, as suggested by an alignment of earthquake hypocentres just above the oceanic Moho (Fig. 2b).

The two seismic belts (Fig. 1) may indicate the lateral extent of the imbricated zone between the E and F reflectors, implying that most of southern Vancouver Island is underlain by a ~110-km-wide duplex structure, an order of magnitude greater than lenses of sediment that can be eroded from the base of accretionary wedges beneath the continental slope²⁸. Within the imbricated zone, a few earthquakes occur where P wave velocities change laterally, suggesting that these earthquakes might be attributable to imbricate thrusting that has juxtaposed different lithologies. If these imbricated rocks are derived mostly from the forearc, then the duplex may result from erosion of the lower crust, perhaps below shelf basins where high inter-plate coupling has been proposed²⁹. Alternatively, if the imbricated rocks originated in the oceanic plate, they would be underplated to the North American plate should aseismic slip along the E reflectors cease, and reorganization of faults within the duplex occur. If slip along the F reflector were to cease, the duplex would become welded to the downgoing slab, and thickened crust would be subducted, as proposed for the Yakutat terrane in Alaska³⁰.

Subduction is usually viewed as occurring along a single, thin interface beneath which an oceanic plate descends into the mantle, and inslab seismicity is considered to be driven by dehydration metamorphism^{3–7}. In the northernmost Cascadia subduction zone, however, the inter-plate boundary is a complex zone up to 16 km in

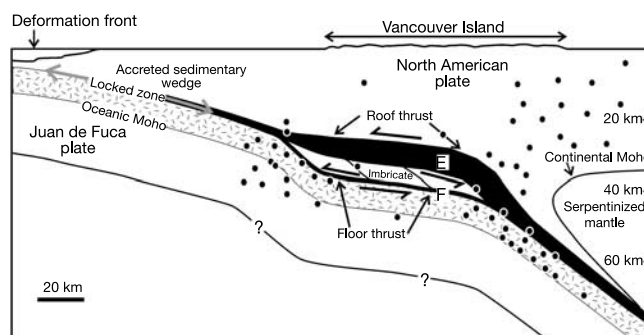


Figure 4 Schematic cross-section of the northern Cascadia convergent margin at southern Vancouver Island. The plate boundary forms a thick duplex structure, within which crustal rocks are imbricated. Inslab earthquakes occur where the dip of the top of the subducting slab, which defines the floor thrust, increases suddenly, suggesting that the earthquakes are caused by deformation in the upper part of the subducting slab as it responds to the geometry of the inter-plate boundary zone. The E reflectors form the roof thrust, and are probably one location of the aseismic, slow-slip events. Slip along the roof thrust will transport the imbricated crustal rocks into the upper mantle above the subducting oceanic slab. Slow slip is also probably occurring along the floor thrust, which is inferred to be the top of the subducting plate. Earthquake hypocentres are indicated by black circles. Vertical exaggeration is 1.5:1.

vertical extent, and inslab seismicity is controlled by deformation in the upper part of the subducting plate. □

Received 11 October 2003; accepted 28 January 2004; doi:10.1038/nature02372.

1. Rogers, G. C. An assessment of the megathrust earthquake potential of the Cascadia subduction zone. *Can. J. Earth Sci.* **26**, 844–852 (1988).
2. Hyndman, R. D. & Wang, K. The rupture zone of Cascadia great earthquakes from current deformation and the thermal regime. *J. Geophys. Res.* **100**, 22133–22154 (1995).
3. Kirby, S. H. *et al.* in *Subduction: Top to Bottom* (ed. Bebout, G. E.) 195–214 (American Geophysical Union, Washington DC, 1996).
4. Peacock, S. M. & Wang, K. Seismic consequences of warm versus cool subduction metamorphism: Examples from southwest and northeast Japan. *Science* **286**, 937–939 (1999).
5. Peacock, S. M. Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* **29**, 299–302 (2001).
6. Hacker, B. R., Peacock, S. M., Abers, G. A. & Holloway, S. D. Subduction factory 2. Are intermediate-depth earthquakes linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108**, doi:10.1029/2001JB001129 (2003).
7. Preston, L. A., Creager, K. C., Crosson, R. S., Brocher, T. M. & Tréhu, A. M. Intraslab earthquakes: Dehydration of the Cascadia slab. *Science* **302**, 1197–1200 (2003).
8. Stanley, D., Villasenor, A. & Benz, H. *Subduction Zone and Crustal Dynamics of Western Washington: A Tectonic Model for Earthquake Hazards Evaluation* Open File Report 99–311 (US Geological Survey, Reston, 1999).
9. Ramachandran, K. *Velocity Structure of S.W. British Columbia and N.W. Washington from 3-D Non-linear Seismic Tomography* Thesis, Univ. Victoria (2001).
10. Spence, G. D., Clowes, R. M. & Ellis, R. M. Seismic structure across the active subduction zone of western Canada. *J. Geophys. Res.* **90**, 6754–6772 (1985).
11. Clowes, R. M., Yorath, C. J. & Hyndman, R. D. Reflection mapping across the convergent margin of western Canada. *Geophys. J. R. Astron. Soc.* **89**, 79–84 (1987).
12. Green, A. G. *et al.* Seismic reflection imaging of the subducting Juan de Fuca plate. *Nature* **319**, 210–213 (1986).
13. Clowes, R. M. *et al.* LITHOPROBE—southern Vancouver Island: Cenozoic subduction complex imaged by deep seismic reflections. *Can. J. Earth Sci.* **24**, 31–51 (1987).
14. Calvert, A. J. & Clowes, R. M. Deep, high-amplitude reflections from a major shear zone above the subducting Juan de Fuca plate. *Geology* **18**, 1091–1094 (1990).
15. Calvert, A. J. Seismic reflection constraints on imbrication and underplating of the northern Cascadia convergent margin. *Can. J. Earth Sci.* **33**, 1294–1307 (1996).
16. Nedimović, M. R., Hyndman, R. D., Ramachandran, K. & Spence, G. D. Reflection signature of seismic and aseismic slip on the northern Cascadia subduction interface. *Nature* **424**, 416–420 (2003).
17. Hasselgren, E. O. & Clowes, R. M. Crustal structure of northern Juan de Fuca plate from multichannel seismic reflection data. *J. Geophys. Res.* **100**, 6469–6486 (1994).
18. Fisher, M. A. *et al.* Seismic survey probes urban earthquake hazards in Pacific Northwest. *Eos* **80**, 16–17 (1999).
19. Brocher, T. M. *et al.* *Wide-angle Seismic Recordings from the 1998 Seismic Hazards Investigation in Puget Sound (SHIPS)*, Western Washington and British Columbia Open File Report 99–314 (US Geological Survey, Reston, 1999).
20. Dehler, S. A. & Clowes, R. M. Integrated geophysical modelling of terranes and other structures along the western Canadian margin. *Can. J. Earth Sci.* **29**, 1492–1508 (1992).
21. Cassidy, J. F. & Ellis, R. M. Shear wave constraints on a deep crustal reflective zone beneath Vancouver Island. *J. Geophys. Res.* **96**, 19843–19851 (1991).

22. Yorath, C. J. *et al.* Marine Multichannel Seismic Reflection, Gravity and Magnetic Profiles—Vancouver Island Continental Margin and Juan De Fuca Ridge (Open File Report 1661, Geological Survey of Canada, Vancouver, 1987).
23. Hyndman, R. D. Dipping reflectors, electrically conductive zones and free water beneath a subduction zone. *J. Geophys. Res.* **93**, 13391–13405 (1988).
24. Dragert, H., Wang, K. & James, T. S. A silent slip event on the deeper Cascadia subduction interface. *Science* **292**, 1525–1528 (2001).
25. Rogers, G. & Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* **300**, 1942–1943 (2003).
26. Kao, H., Shan, S.-J., Rogers, G., Cassidy, J. F. & Dragert, H. Temporal and spatial distribution of the tremors in the episodic tremor and slip (ETS) event observed beneath the northern Cascadia subduction zone in early 2003. *Eos* **84** (Fall Meeting Suppl.), S42G–02 (2003).
27. Bolton, M. K. & Rogers, G. C. Juan de Fuca plate seismicity at the northern end of the Cascadia subduction zone. *Seismol. Res. Lett.* **73**, 214 (2002).
28. Ranero, C. R. & von Huene, R. Subduction erosion along the Middle America convergent margin. *Nature* **404**, 748–752 (2003).
29. Wells, R. E., Blakely, R. J., Sugiyama, Y., Scholl, D. W. & Dinterman, D. A. Basin-centered asperities in great subduction zone earthquakes: A link between slip, subduction and subduction erosion. *J. Geophys. Res.* **108**, doi:10.1029/2002JB002072 (2003).
30. Brocher, T. M., Fuis, G. S., Fisher, M. A., Plafker, G. & Moses, M. J. Mapping the megathrust beneath the northern Gulf of Alaska using wide-angle seismic data. *J. Geophys. Res.* **99**, 11663–11685 (1994).

Acknowledgements I thank K. Ramachandran for making available to me the 3D velocity model and the positions of the relocated earthquakes; T. Brocher for comments and suggestions that helped to improve the final manuscript; and H. Kao and M. Nedimović for discussions. This project was funded by the Natural Sciences and Engineering Research Council of Canada.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (acalvert@sfu.ca).

Unified spatial scaling of species and their trophic interactions

Ulrich Brose^{1,2*}, Annette Ostling³, Kateri Harrison² & Neo D. Martinez²

¹Department of Computer Science, Thorton Hall 906, 1600 Holloway Avenue, San Francisco State University, San Francisco, California 94132-4163, USA

²Pacific Ecoinformatics and Computational Ecology Laboratory, Rocky Mountain Biological Laboratory, POB 519, Gothic, Colorado 81224, USA

³Energy and Resources Group, University of California, Berkeley, California 94720-3050, USA

* Present address: Technical University of Darmstadt, Department of Biology, Schnittspahnstr. 3, D-64287, Darmstadt, Germany

Two largely independent bodies of scaling theory address the quantitative relationships between habitat area, species diversity and trophic interactions. Spatial theory within macroecology addresses how species richness scales with area in landscapes, while typically ignoring interspecific interactions^{1–6}. Complexity theory within community ecology addresses how trophic links scale with species richness in food webs, while typically ignoring spatial considerations^{7–12}. Recent studies suggest unifying these theories by demonstrating how spatial patterns influence food-web structure^{13–16} and vice versa¹⁷. Here, we follow this suggestion by developing and empirically testing a more unified scaling theory. On the basis of power-law species–area relationships, we develop link–area and non-power-law link–species models that accurately predict how trophic links scale with area and species richness of microcosms, lakes and streams from community to metacommunity levels. In contrast to previous models that assume that species richness alone determines the number of trophic links^{7,8}, these models include the species’ spatial distribution, and hence extend the domain of complexity theory to metacommunity scales. This generality and predictive success shows how complexity theory and spatial theory can be unified into a much more general theory addressing new domains of ecology.

A widely observed pattern in ecology is the power-law scaling of the number of species S of a metacommunity with local habitat area A

$$S = cA^z \quad (1)$$

where c and z are positive constants^{1,2}. This relationship may occur when colonization frequency increases with area and extinctions become less likely in larger habitats because of larger populations^{1–3}, in which case spatial isolation can reduce z by preventing colonization events^{1–3}. A recent incorporation of trophic considerations into spatial scaling theory suggests that z increases with trophic level¹⁷. Here, we reinforce this bridge between fields in the opposite direction by incorporating spatial scaling considerations into trophic theory.

Theories of biocomplexity within local communities express the power-law scaling of feeding links between species, L , with species richness, S , as

$$L = bS^u \quad (2)$$

where b and u are positive constants^{7–10}. The ‘link–species scaling law’ asserts that consumers have a finite number of resources that is independent of community diversity ($u = 1$)⁷. The alternative ‘constant connectance hypothesis’ asserts that species are linked on average to a fixed fraction of other species, and that links scale with the square of species richness ($u = 2$)⁸. Empirical analyses find exponents spanning this range ($1 \leq u \leq 2$)^{7–10} and classical stability theory predicts exponents at the low end of this range due to positive feedbacks thought to accompany higher linkage densities¹⁸. However, spatial patterns may ameliorate this instability of highly diverse food webs^{16,18–20}. A few studies have found that ecosystem size determines the length of food chains^{14–16}, but such study of the effects of space on food-web structure is surprisingly limited. Here, we address this limitation by deriving link–area and spatially explicit link–species relationships from species–area models and testing them against the best available data.

Combining equations (1) and (2) yields a simple model for the scaling of links with area

$$L = bc^u A^{uz} \quad (3)$$

This link–area relationship depends on whether the link–species scaling law ($u = 1$)⁷ or the constant connectance hypothesis ($u = 2$)⁸ is accepted. Being based on equation (2), equation (3) also assumes that the link–species relationship in equation (2) holds independent of the value of S . This characterizes local communities, in which all S species co-occur spatially and therefore all consumers co-occur with and consume their resources among the S species⁸. In metacommunities, species may be spatially isolated from potential consumers and resources, which might prevent certain links from being realized⁸. This suggests that equations (2) and (3) inaccurately predict trophic complexity at macroecological scales^{8,21}. Here, we address this potential inaccuracy by starting with the species–area model in equation (1) and deriving the following food-web complexity models, which do not assume general consumer–resource co-occurrence (see Box 1):

$$L = K\theta(A)A^{2z} = (K/c^2)\theta(A)S^2 \quad (4)$$

$$L = K\theta(A)A^{z_k+z_r} = \frac{K}{c_k c_r} \theta(A) S_k S_r = \frac{K}{c_k c_r} \theta(A) S^u \quad (5)$$

where K is constant. In contrast to equation (4), equation (5) accounts for differences in the species–area relationships between consumers and resources, as indicated by the subscripts k and r , respectively. $\theta(A)$ represents the increased or decreased likelihood of consumer species occurrences in patches with their resources compared with random patches. This effectiveness of consumers tracking their resources may be scale dependent. For the metacommunity in area A_0 , $L(A_0) = L_0$ if $\theta(A_0) = 1$. At other spatial scales, $\theta(A) > 1$ or $\theta(A) < 1$ indicates that, on average, consumer species

have a higher or lower than random probability of co-occurring with their resources in areas of size A , respectively. The exponent u of the link–species relationship is

$$u = \ln(S_k S_r) / \ln(S) \quad (6)$$

Box 1

The unified spatial scaling model

We consider a real or fictitious habitat patch of area A_0 large enough for a metacommunity of S_0 species (according to equation (1)) with L_0 trophic links, and local community patches of area $A_i = A_0/2^i$ containing S_i species on average. The species–area relationship of equation (1) implies^{5,6} that a randomly chosen species will be present in a random local community patch of size A_i with the probability

$$p_i = S_i/S_0 = a^i \quad (7)$$

where $a = 2^{-z}$. Equation (7) exhibits the fractal or self-similarity property^{5,6} $p_i = ap_{i-1}$ for all i . The central tendency among species, p_i , does not necessarily describe each individual species. The mean number of links L_i in A_i can be calculated by

$$L_i = \sum_{k=1}^{S_0} \sum_{r \rightarrow k} p_i(k|r) p_i(r) \quad (8)$$

where $p_i(r)$ is the patch occupancy of resource species r , $p_i(k|r)$ is the occurrence probability of consumer species k in patches where its resource species r is present, and $\sum_{r \rightarrow k}$ is a sum over all species that are eaten by species k . As in empirical studies, equation (8) assumes that species trophically linked in the metacommunity will be linked in any local community where both species are present. If z and c are independent of species' trophic rank, equation (8) yields equation (4) (see main text for equation (4) and see Supplementary Information for assumptions and details of derivation), where $K = L_0/A_0^{2z}$ and $\theta(A)$ represents the dependence of consumer on resource occurrence and is defined as $\theta(A_i) = \langle p_i(k|r)/p_i(r) \rangle_{r \rightarrow k}$. Accounting for differences in z between resource and consumer species, equation (8) yields equation (5) (see main text), where k and r label properties of the consumers and resources, respectively, and S_t , z_t and c_t are the species richness, species–area exponent and species–area constant for trophic level t , and $K = L_0/(A_{0,k}^{2z_k} A_{0,r}^{2z_r})$, where $A_{0,k}$ and $A_{0,r}$ are the area sizes large enough to hold all of the consumer and resource species in the metacommunity, respectively. At A_0 , equation (5) correctly gives $L(A_0) = L_0$, as long as $A_{0,k} = A_{0,r} = A_0$ and $\theta(A_0) = 1$. The exponent of the link–species relationship is given in equation (6) (see main text).

If S_k , S_r and S all follow power-law species–area relationships, one might conclude $u = (z_k + z_r)/z$ (see Supplementary Information). However, $S = d(A)(c_k A^{z_k} + c_r A^{z_r})$, where $d(A)$ is a scale-dependent fraction that accounts for overlap among species groups. As the sum of power laws with different exponents is not a power law, S_k , S_r and S cannot all follow power-law species–area relationships and u cannot be constant. This contrasts with equation (4), where u may be constant. Such constancy occurs in unlikely metacommunities having neither basal (resources without resources) nor top (consumers without consumers) species. This causes $S_k = S_r = S$ and $u = 2$ at all scales. u cannot exceed 2 over a large range⁸ because L cannot exceed the square of S . In equation (5), u decreases from 2 with increasing diversity of basal and top species (equation (6)). The extreme $u \approx 1$ occurs in metacommunities with all but one species being either all basal or all top species. Scale-dependent co-occurrence factors ($\theta(A)$) can also prevent metacommunity link–species relationships (equations (4) and (5)) from being power laws. These examples show that our link–species model of equation (5) includes previous models asserting $u = 1$ and $u = 2$ as extremes and allows for recently observed^{7–10} intermediate values of u . L may scale with $S_k S_r$ even across food webs belonging to different metacommunities, because L cannot exceed $S_k S_r$.

Unlike equations (2) and (3), equations (4) and (5) use metacommunity food-web structure and the degree of co-occurrence of consumers with their resources to predict the scaling of links with habitat area and species richness, enabling them to be applied at local to macroecological scales. Equations (2) and (3) are 'community scaling models' because they assume general consumer–resource co-occurrence as happens in local communities. Equations (4) and (5) are 'unified scaling models' because they integrate species–area models and scale-dependent consumer–resource co-occurrences to predict how complexity scales from local communities to macroecological scales.

We tested the community and unified models against empirical data from aquatic microcosms²², lakes^{23,24} and streams^{25–27}. The microcosms include a mean of 203 trophic links among 32 taxa with only one primary producer and one top consumer (see Methods). This causes consumer and resource species to be very similar with expectedly similar z values, allowing the use of equation (4). The empirically determined link–area exponent equals 2.19 z (Fig. 1a), which is reasonably close to our predicted value of $2z$ in equations (4) or (3) under constant connectance, and substantially higher than the previously suggested z (ref. 7).

The larger and more complex lake and stream metacommunities (Table 1) also have very similar z values for consumers and resources

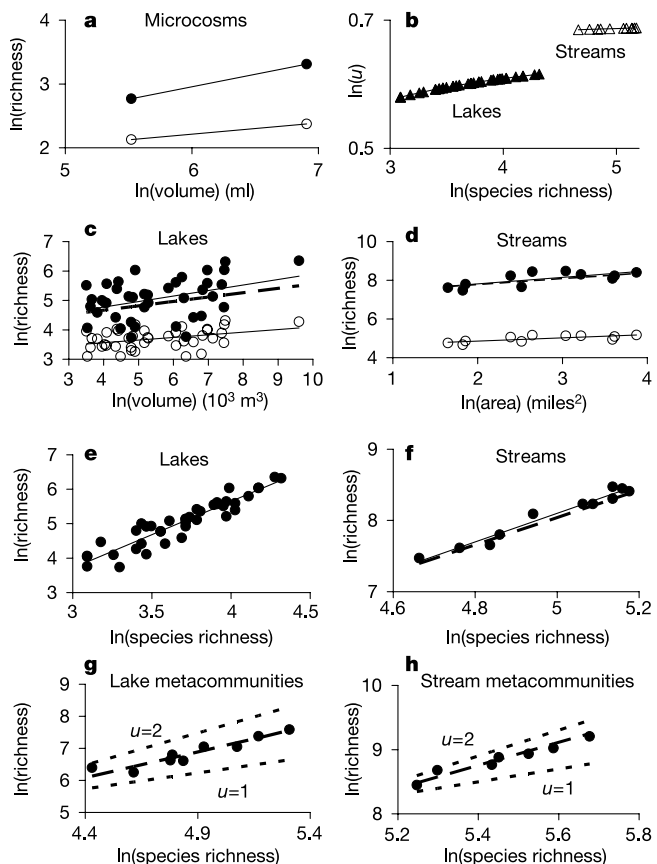


Figure 1 Scaling of links (filled circles) and species (open circles) in aquatic microcosms, lakes and streams with fitted regressions (solid lines), the unified scaling model (equation (5); dashed lines) and prior link–species models (dotted lines). Note that predictions from the model of equation (4) are not shown as they are indistinguishable from equation (5) in most panels. Regression and model fit details for the lakes and streams are given in Tables 2 and 3. Other lines fitted to data are $S = 3.15A^{0.18}$ and $L = 1.85A^{0.39}$ for microcosms, $u = \ln(0.53S^2 - 0.32S - 0.25)/\ln(S)$ for lakes, and $u = \ln(0.98S^2 - 5.26S - 5.86)/\ln(S)$ for streams.

Table 1 Community characteristics of the Adirondack lakes^{23,24} and Santa Clara Valley streams^{25–27}

Data	A_0	Local A	S_0	$S_{0,K}$	$S_{0,r}$	Local S	L_0	Local L	Range of u	Mean θ
AL	1.23×10^{10}	20–14,778	202	96	198	13–75	1,975	17–571	1.73–1.85	1.8
SC	978	5.2–48.1	292	286	286	106–177	9,940	1,758–4,777	1.98–1.99	1.3

(Table 2). The co-occurrence factors of these consumers and resources, θ , as a function of area were closely fitted by power laws (Table 2). Mean θ -values of 1.8 and 1.3 for the local communities indicate higher than random co-occurrence of consumers and resources (Table 1). We used power-law fits to the scale-dependent θ -values (Table 2) to empirically test our unified models (equations (4) and (5)). The link–area relationships predicted by equations (4) and (5) are extremely close to the plotted regression lines for the lakes (Fig. 1c) and are visually indistinguishable for the streams (Fig. 1d). The calculated exponent u varied across the species richness range of the lakes (Table 2), as shown by the curved line that indicates a non-power-law relationship with S (Fig. 1b). In the more species-rich stream food webs, however, there was very little variance in u (Table 2), which increases with S very slightly (Fig. 1b). If variation in u generally decreases as S increases, then power-law link–species relationships would fit much better in metacommunities with large food webs than in those with small food webs¹⁰. The link–species relationship based on the scale-dependent u -values (equation (5)) provides an excellent fit to the lake and stream data sets (Fig. 1e, f and Table 2). However, similar fits are obtained with

equation (4) and with constant u -values depending on the z -values (Table 3; see also Supplementary Information).

Table 3 compares the fit of the models of equations (2) and (3) with $u = 1$ or $u = 2$ and the unified scaling models of equations (4) and (5) to the more detailed lake and stream data. The link–species scaling law ($u = 1$) poorly fits the link–species data, whereas the constant connectance ($u = 2$) and our unified scaling models fit quite well ($r^2 > 0.84$). Using $u = 1$ in the community model also poorly fits link–area data, whereas all other models provide fits equal to the underlying species–area relationships (Tables 2 and 3). Extending the predictions to macroecological data that combine adjacent local food webs (see Methods and Supplementary Information) distinguishes the models much more clearly. The community models markedly overestimate (constant connectance) or underestimate (link–species scaling law) the number of links, whereas the unified models of equations (4) and (5) yield accurate estimates (Fig. 1g, h). The macroecological fits of the unified models were as high as the local fits ($r^2 > 0.84$), whereas the community models yielded poorer fits at this scale (Table 3).

Empirical comparison of these models is confounded by differences in the number of fit parameters. One typically expects better fits for the unified scaling models with additional parameters, but the simpler constant connectance and community link–area models provided similar fits to the data of the local communities here. This is probably due to the very similar z -values for consumers and resources in our empirical data, which forces the unified model of equation (5) to have similar link–species ($u = [z_k + z_r]/z \approx 2$) and link–area ($z_k + z_r \approx 2z$) exponents as the community models of equations (2) and (3). However, other recent studies⁹ find $u \approx 1.3$, suggesting different z -values for consumers and resources or macroecological study scales, in which cases the unified model of equation (5) should fit much better than the community model with $u = 2$. Data with z_k very different from z_r are needed to better test the community-level predictions of the unified and community models. Still, the major predictive contribution of our unified scaling models is their accuracy from local to macroecological scales. Being based on the metacommunity perspective they provide very close fits to macroecological data. At local scales, the scale-dependent co-occurrence factor ($\theta(A) > 1$) allows for higher than random consumer–resource co-occurrence probabilities and yields empirical fits of similar quality as the community models. Community scaling models that ignore co-occurrence probabilities seem

Table 2 Fitted regressions for the Adirondack lakes and Santa Clara Valley streams

Data	Dependent	Independent	Fitted power laws			
			Constant	Exponent	r^2	P
AL	Species	Area	24.5	0.09	0.152	0.01
AL	Consumers	Area	12.4	0.093	0.118	0.03
AL	Resources	Area	24.5	0.09	0.152	0.01
SC	Species	Area	91.8	0.168	0.537	0.01
SC	Consumers	Area	86.2	0.176	0.537	0.01
SC	Resources	Area	90.6	0.167	0.534	0.01
AL	θ	Area	2.2	−0.035	0.812	0.01
SC	θ	Area	1.4	−0.040	0.744	0.01
AL	Links	Area	55.7	0.187	0.155	0.01
SC	Links	Area	1,256	0.339	0.522	0.01
AL	Links	Species	0.125	1.93	0.872	0.001
SC	Links	Species	0.148	2.00	0.964	0.001

Data	Dependent	Independent	Fitted linear regressions			
			Slope	Intercept	r^2	P
AL	Consumers	Species	0.52	−0.25	0.813	0.001
AL	Resources	Species	1	−0.13	0.999	0.001
SC	Consumers	Species	1	−6	1	0.001
SC	Resources	Species	0.98	0.6	0.998	0.001

Table 3 Comparison of model fits (r^2) to empirical link–species and link–area data

Data	Equation (5)	Equation (4)	Link–species fits		
			Constant connectance	Link–species scaling law	Fitted power law
AL	0.868	0.849	0.849	0.316	0.872
SC	0.929	0.940	0.954	0.482	0.964

Data	Equation (5)	Equation (4)	Link–area fits		
			Equation (3) ($u = 2$)	Equation (3) ($u = 1$)	Fitted power law
AL	0.146	0.120	0.145	0.012	0.155
SC	0.466	0.452	0.518	0.175	0.522

Data	Equation (5)	Equation (4)	Macroecological link–species fits		
			Constant connectance	Link–species scaling law	Fitted power law
AL	0.844	0.900	0.620	0.515	0.648
SC	0.948	0.947	0.753	0.528	0.767

unlikely to survive tests at such highly divergent scales.

Separate trophic and spatial scaling models focus on smaller and larger spatial scales, respectively. Trophic scaling theories that ignore the relative spatial distribution of consumer and resource species are expected to break down at large scales^{8,21}. Here, we combine the species–area scaling theory of biogeography and the link–species scaling theory of food webs. The generation and testing of unified scaling theory thus formalizes and helps to illuminate only recently explored spatial aspects of food-web theory¹⁶. Our unified models extend the predictive domain of trophic scaling theory to macroecological scales. Moreover, habitat areas can now be used to estimate species richness and connectance that can parameterize the niche model, which may successfully predict the detailed structure of the community food webs¹¹. Without unification, species–area and link–species theories are far less general and predictive.

Our results show that scale-dependent co-occurrence factors yield accurate predictions across scales from local communities to metacommunities. This challenges spatial community-assembly theories that assume colonization and extinction dynamics are random with respect to species identity^{2–5}. Although such randomness may take place at larger spatial scales⁵, locally nonrandom co-occurrence ($\theta > 1$) suggests that consumer species spatially track their trophic resources or vice versa; for example, in the case of plants that require pollinator species. Such nonrandom community assembly with respect to species' trophic identity should be incorporated into spatial assembly models and empirically tested⁶. Studies of the magnitude of this co-occurrence factor in different habitat types at different scales could make significant contributions to community-assembly theory.

The unified scaling model presented here connects food-web and spatial ecology²⁸ and could help to articulate the role of spatial processes in determining food-web structure¹⁶. This connection offers new possibilities for theoretical community models as well as for conservation ecology by providing a framework for exploring spatial aspects of interspecific interactions that are critical to species' survival. In particular, we hope that such unified understanding of community processes may help to address some of ecology's most challenging issues, including the stability of diverse and complex natural food webs^{16,18–20} and community responses to biodiversity loss²⁹. □

Methods

Species richness and the number of trophic links were analysed for Adirondack mountain lakes^{23,24}, Santa Clara Valley streams^{25–27} and aquatic microcosms²². In the microcosm study²², glass beakers of 250 ml and 1,000 ml, with 12 replicates of each size, were incubated with the same initial species composition. The microcosms were sampled after 163 days to measure species richness and the number of trophic links. The effects of habitat volume on species richness ($P = 0.005$) and the number of links per species ($P = 0.007$) were significant²². The mean numbers of species and links for each size class were analysed for species–area and link–area relationships. Tests were restricted to the model of equation (4) because the specific food webs in each beaker were not available. Such specific food-web data were available in the lake and stream data sets. Feeding links were based on literature surveys and assumed present when both species were found in the same habitat (see refs 23–27 for details). This less labour-intensive convention typically results in a greater number of links than more direct observations of feeding within the habitats studied⁹, which typically yields a minimum number of links unless sampling is exhaustive³⁰. Lake volume (10^3 m^3) and watershed area (miles^2) estimated ecosystem size in the Adirondack lakes and Santa Clara Valley streams, respectively. Watershed area is an appropriate measure of stream habitat area because watershed area was highly (Pearson's $r = 0.86$) and significantly ($P < 0.001$) correlated with the length from the stream's mouth to its most upstream sampling station.

Of the 50 lakes, we excluded 10 that contained both high monomeric aluminium ($>15 \mu\text{equiv. l}^{-1}$) and low dissolved organic carbon ($<7 \text{ mg l}^{-1}$). Such toxic aluminium reduces species richness, and low carbon indicates a reduced ability of humic materials to ameliorate this effect by complexing aluminium²³. The remaining 40 lakes represent samples of different sizes from the same metacommunity²³. Of the 13 streams^{25–27}, we excluded two strong outliers with extremely large watershed areas and comparatively low species richness. The average co-occurrence factors, θ , across predator and prey combinations were calculated for seven different-sized groups of lakes and four different-sized groups of streams. Linear least-squares regressions in log–log data space analysed species–area, link–area, link–species and θ –area relationships. We calculated the

dependence of u on S using linear regressions of the number of consumer and the number of resource species versus total species richness, S .

The metacommunities were generated by combining communities in each study area. All communities were combined to generate the largest-scale data point. Each study area was then divided into subareas for which species richness and realized links within all food webs included were counted. The Adirondack lakes were divided into a southern (south of $43^\circ 55' \text{ N}$) and a northern part. Six additional smaller metacommunities were yielded by dividing the southern (north of $43^\circ 45' \text{ N}$, north of $43^\circ 35' \text{ N}$, south of $43^\circ 35' \text{ N}$) and northern subareas (west of $74^\circ 30' \text{ E}$, west of $74^\circ 15' \text{ E}$, east of $74^\circ 15' \text{ E}$). The Santa Clara Valley streams were divided into those coming from the eastern and western slopes. Four additional metacommunities were represented by the San Francisco Creek system, the Guadalupe River system, the Coyote Creek system and the Saratoga and Stevens Creek system. These systems included all of their tributaries. This resulted in nine and seven macroecological data points for the Adirondack lakes and Santa Clara Valley streams, respectively. We compared the predictions of the link–species models (equations (2), (4) and (5)) at the macroecological S_i . The locally fitted co-occurrence factor $\theta(A)$ was used in equations (4) and (5), and this power law was evaluated at the area A_i , which would hold all of the S_i species according to the observed total species–area relationship for each community. In addition, the dependence of u on S obtained from the local scale fits was extrapolated to these macroecological scales.

Received 5 November; accepted 16 December 2003; doi:10.1038/nature02297.

- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
- Connor, E. F. & McCoy, E. F. The statistics and biology of the species–area relationship. *Am. Nat.* **113**, 791–833 (1979).
- MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, Princeton, New Jersey, 1962).
- Hubbell, S. P. *A Unified Theory of Biogeography and Relative Species Abundance* (Princeton Univ. Press, Princeton, New Jersey, 1997).
- Harte, J. & Kinzig, A. P. On the implications of species–area relationships for endemism, spatial turnover, and food web patterns. *Oikos* **80**, 417–427 (1997).
- Green, J., Harte, J. & Ostling, A. Species richness, endemism and abundance patterns: tests of two fractal models in a serpentine grassland. *Ecol. Lett.* **6**, 919–928 (2003).
- Cohen, J. E. & Briand, F. Trophic links of community food webs. *Proc. Natl Acad. Sci. USA* **81**, 4105–4109 (1984).
- Martinez, N. D. Constant connectance in community food webs. *Am. Nat.* **139**, 1208–1218 (1992).
- Schmid-Araya, J. M. *et al.* Connectance in stream food webs. *J. Anim. Ecol.* **71**, 1056–1062 (2002).
- Bersier, L. F. & Sugihara, G. Scaling regions for food web properties. *Proc. Natl Acad. Sci. USA* **94**, 1247–1251 (1997).
- Williams, R. J. & Martinez, N. D. Simple rules yield complex food web dynamics. *Nature* **404**, 180–183 (2000).
- Pimm, S. L., Lawton, J. H. & Cohen, J. E. Food web patterns and their consequences. *Nature* **350**, 669–674 (1991).
- Schoener, T. W. Food webs from the small to the large. *Ecology* **70**, 1559–1589 (1989).
- Cohen, J. E. & Newman, C. M. Community area and food chain length—theoretical predictions. *Am. Nat.* **138**, 1542–1554 (1991).
- Post, D., Pace, M. & Hairston, N. G. Ecosystem size determines food chain length in lakes. *Nature* **405**, 1047–1049 (1999).
- Holt, R. D. Food webs in space: On the interplay of dynamic instability and spatial processes. *Ecol. Res.* **17**, 261–273 (2002).
- Holt, R. D., Lawton, J. H., Polis, G. A. & Martinez, N. D. Trophic rank and the species–area relationship. *Ecology* **80**, 1495–1504 (1999).
- May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, Princeton, New Jersey, 1974).
- Brose, U., Williams, R. J. & Martinez, N. D. Comment on “Foraging adaptation and the relationship between food-web complexity and stability”. *Science* **301**, 918 (2003).
- Holyoak, M. Habitat patch arrangement and metapopulation persistence of predator and prey. *Am. Nat.* **156**, 378–389 (2000).
- Martinez, N. D. & Lawton, J. H. Scale and food-web structure—from local to global. *Oikos* **73**, 148–154 (1995).
- Spencer, M. & Warren, P. H. The effects of habitat size and productivity on food web structure in small aquatic microcosms. *Oikos* **75**, 419–430 (1996).
- Sutherland, J. W. (ed.) *Field Survey of the Biota and Selected Water Chemistry Parameters in 50 Adirondack Mountain Lakes* (DEC Publication, New York, 1989).
- Havens, K. Scale and structure in natural food webs. *Science* **257**, 1107–1109 (1992).
- Harrison, K. *Effects of Land Use and Dams on Stream Food Web Ecology in Santa Clara Valley, California* 127. Thesis, San Francisco State Univ. (2003).
- Carter, J. L. & Fend, S. V. *The Distribution and Abundance of Lotic Macroinvertebrates During Spring 1998 in Seven Streams of the Western Santa Clara Valley Area* (California US Geological Survey, Open-File Report 00–346, 2000).
- Carter, J. L. & Fend, S. V. *The Distribution and Abundance of Lotic Macroinvertebrates During Spring 1998 in Seven Streams of the Santa Clara Valley Area* (California US Geological Survey, Open-File Report 00–68, 2000).
- Martinez, N. D. in *Linking Species and Ecosystems* (eds Jones, C. G. & Lawton, J. H.) 166–175 (Chapman and Hall, New York, 1995).
- Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
- Martinez, N. D., Hawkins, B. A., Dawah, H. A. & Feifarek, B. P. Effects of sampling effort on characterization of food-web structure. *Ecology* **80**, 1044–1055 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Harte and J. Green for comments on the manuscript. U.B. is supported by the German Academy of Naturalists Leopoldina and by funds from the German Federal Ministry of Education and Science. A.O. is supported by the National Center for Environmental Research (NCER) STAR Program, EPA. N.D.M. is supported by the Cornell Telluride Association, Cornell Center for Applied Mathematics, and the US National Science Foundation Biocomplexity, Biological Databases and Informatics, Information Technology Research, and Interdisciplinary Graduate Education and Research Training programs.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.B. (brose@sfsu.edu).

Pervasive alteration of tree communities in undisturbed Amazonian forests

William F. Laurance^{1,2}, Alexandre A. Oliveira⁴, Susan G. Laurance^{1,2}, Richard Condit¹, Henrique E. M. Nascimento¹, Ana C. Sanchez-Thorin¹, Thomas E. Lovejoy², Ana Andrade², Sammya D'Angelo², José E. Ribeiro³ & Christopher W. Dick¹

¹Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Republic of Panamá

²Biological Dynamics of Forest Fragments Project, and ³Department of Botany, National Institute for Amazonian Research (INPA), CP 478, Manaus, AM 69011-970, Brazil

⁴Department of Biology, University of São Paulo, Avenida Bandeirantes 3900, Ribeirão Preto-São Paulo, SP 14040-901, Brazil

Amazonian rainforests are some of the most species-rich tree communities on earth¹. Here we show that, over the past two decades, forests in a central Amazonian landscape have experienced highly nonrandom changes in dynamics and composition. Our analyses are based on a network of 18 permanent plots unaffected by any detectable disturbance. Within these plots, rates of tree mortality, recruitment and growth have increased over time. Of 115 relatively abundant tree genera, 27 changed significantly in population density or basal area—a value nearly 14 times greater than that expected by chance. An independent, eight-year study in nearby forests corroborates these shifts in composition. Contrary to recent predictions^{2–5}, we observed no increase in pioneer trees. However, genera of faster-growing trees, including many canopy and emergent species, are increasing in dominance or density, whereas genera of slower-growing trees, including many subcanopy species, are declining. Rising atmospheric CO₂ concentrations⁶ may explain these changes, although the effects of this and other large-scale environmental alterations remain uncertain. These compositional changes could have important impacts on the carbon storage, dynamics and biota of Amazonian forests.

Recent studies suggest that undisturbed Amazonian forests are becoming more dynamic over time, with higher rates of tree mortality and turnover^{4,5}, and that carbon storage^{7–10} and productivity¹¹ in these forests are increasing. In addition, lianas—climbing woody vines that often thrive in disturbed forest—appear to be increasing in size and abundance¹². A plausible cause of these changes is increasing plant fertilization caused by rising atmospheric CO₂ concentrations^{2–13}, although other large-scale phenomena, such as alterations in regional temperature¹⁴, rainfall^{15,16}, available solar radiation¹⁷, or nutrient deposition¹⁸, might also play a role. However, no studies have assessed the effects of such

large-scale changes on the taxonomic composition of Amazonian tree communities, which greatly influences the architecture, dynamics and ecological functioning of these forests.

We assessed long-term changes in tree-community composition within a network of 18 discrete, one-hectare plots in central Amazonia (Supplementary Figs 1 and 2). The plots span an area of about 300 km², are randomly located with respect to local topography, and are positioned at least 300 m away from any clearings to avoid edge effects^{19,20}. Plots exhibited no evidence of current or past disturbance from logging, fires, hunting or major windstorms, although two plots experienced small wet-season floods that caused temporary increases in tree mortality. All plots were established from 1981 to 1987 and re-censused at roughly five-year intervals for an average of 15.0 yr (range = 11.4 to 18.2 yr), with the final census of each in 1999 or 2000. Within each plot, all trees (≥10 cm diameter at breast height, DBH) were marked with permanent tags, mapped, measured for trunk diameter, and identified on the basis of sterile or fertile material. In total, nearly 13,700 trees were recorded.

We assessed changes in the abundance of tree genera, rather than species, for three reasons. First, 88% of tree species in our study area are too rare (<1 individual per hectare) to allow robust analyses of population trends. Second, congeneric species of Amazonian trees tend to be similar ecologically^{21,22}, so analyses at the genus level capture most of the relevant information. Third, 95.3% of study trees were positively identified to genus, whereas a smaller percentage was identified to species.

We encountered 244 tree genera in our plots, of which 115 were sufficiently abundant (initially present in at least 8 of 18 plots) to permit rigorous analysis. For each genus, we used bootstrapping (see Methods) to test for changes in population density and basal area (a strong correlate of tree biomass) between the first and final

Table 1 Increasing or decreasing tree genera in undisturbed Amazonian rainforests

Genus	Family	Net change (%)
Tree density increases over time		
<i>Corythophora</i>	Lecythidaceae	+9.8
<i>Eschweilera</i>	Lecythidaceae	+4.0
Tree density decreases over time		
<i>Aspidosperma</i>	Apocynaceae	–13.3
<i>Brosimum</i>	Moraceae	–8.1
<i>Couepia</i>	Chrysobalanaceae	–8.9
<i>Croton</i>	Euphorbiaceae	–35.0
<i>Heisteria</i>	Oleaceae	–25.0
<i>Hirtella</i>	Chrysobalanaceae	–13.0
<i>Iryanthera</i>	Myristicaceae	–16.3
<i>Licania</i>	Chrysobalanaceae	–11.0
<i>Nucleopsis</i>	Moraceae	–17.8
<i>Oenocarpus</i>	Arecaceae	–32.3
<i>Quilina</i>	Quinaceae	–29.0
<i>Tetragastris</i>	Burseraceae	–15.0
<i>Unonopsis</i>	Annonaceae	–15.3
<i>Virola</i>	Myristicaceae	–14.0
Tree basal area increases over time		
<i>Corythophora</i>	Lecythidaceae	+12.0
<i>Couepia</i>	Chrysobalanaceae	+10.8
<i>Couma</i>	Apocynaceae	+14.4
<i>Dipteryx</i>	Leguminosae	+7.2
<i>Ecclinusa</i>	Sapotaceae	+13.8
<i>Eschweilera</i>	Lecythidaceae	+7.0
<i>Licaria</i>	Lauraceae	+17.2
<i>Maquira</i>	Moraceae	+9.9
<i>Parkia</i>	Leguminosae	+22.0
<i>Peltogyne</i>	Leguminosae	+15.9
<i>Sarcocaulis</i>	Sapotaceae	+14.4
<i>Sclerobium</i>	Leguminosae	+76.6
<i>Sterculia</i>	Sterculiaceae	+23.4
<i>Trattinnickia</i>	Burseraceae	+13.6
Tree basal area decreases over time		
<i>Oenocarpus</i>	Arecaceae	–29.1

All increases or decreases in tree genera based on population density and basal-area data are significant ($P < 0.01$).

censuses of the plots. Using a conservative 1% significance level in our tests, we expected—for each parameter—that about one out of 115 genera would show a significant change by chance alone.

Overall, we found 31 significant changes, with 14 genera increasing in basal area and 14 genera declining in density (Table 1). One genus, *Couepia*, simultaneously increased in basal area while declining in density (the result of increased tree growth but high mortality of small individuals), whereas three other genera either decreased (*Oenocarpus*) or increased (*Corythophora*, *Eschweilera*) in both density and basal area. Thus, excluding *Couepia*, 13 genera declined in density, and 13 genera increased in basal area, sometimes dramatically (Table 1). Most genera that declined in density did not also decline in basal area because of the accelerated growth of the surviving trees (see below).

Mortality rates differed between the 13 increasing and 13 decreasing genera. Decreasing genera had much higher mortality than increasing genera (1.57 ± 0.90 versus $0.51 \pm 0.31\%$ yr^{-1} ; $t = 4.66$, degrees of freedom, d.f. = 24, $P = 0.0001$; t -test with log-transformed data), whereas recruitment rates did not differ between the two (0.50 ± 0.48 versus $0.69 \pm 0.42\%$ yr^{-1} ; $t = 1.06$, d.f. = 24, $P = 0.30$; t -test). Recruitment rates of increasing and decreasing genera were both on average lower than the stand-level rate (1.06% yr^{-1}) because they included few pioneers (which have higher recruitment).

These shifts in tree communities were not driven by large overall

changes in tree density or basal area. During the course of our study, average tree density declined by 1.1%, whereas average basal area rose by 1.9%. Neither change was statistically significant ($P > 0.09$; paired t -tests).

Two lines of evidence confirm that these compositional changes reflect underlying biological processes, not sampling errors. First, randomization tests revealed that the observed changes in density ($P = 0.001$) and basal area ($P = 0.002$) for all 115 genera were consistent across the 18 plots (see Methods). Second, we contrasted our results with those of a separate, eight-year study¹, in which trees in three one-hectare plots were censused in undisturbed forest about 6 km east of our study area (see Methods). Changes over time in both density and basal area of genera were significantly and positively correlated between the two studies (Fig. 1). Thus, parallel studies by two separate teams of investigators revealed similar patterns of change.

Do the increasing and declining tree genera differ biologically? We reviewed available literature and Internet resources and used data from our long-term study to quantify key ecological traits for most genera (Supplementary Table 1). The 13 increasing genera and 13 decreasing genera differed in growth form: all of the former were canopy or emergent trees, whereas six (46%) of the latter were subcanopy trees (the remainder being canopy or emergent trees), a highly significant difference ($G = 10.15$, d.f. = 1, $P = 0.001$; G -test). Similarly, among all 115 genera, there was a clear tendency for large trees to increase in population density (Fig. 2) and basal area at the expense of small trees.

Surprisingly, successional status differed little between the 13 increasing and 13 declining genera; old-growth trees dominated (77%) both groups. In addition, none of the major pioneer genera (*Annona*, *Cecropia*, *Croton*, *Goupia*, *Jacaranda*, *Miconia*, *Pourouma*, *Vismia*) increased significantly in density or basal area, either individually or when pooled (Supplementary Table 2). Nevertheless, both median ($t = 2.28$, d.f. = 24, $P = 0.032$) and maximum ($t = 2.07$, d.f. = 24, $P = 0.049$) absolute growth rates were significantly higher in the increasing than declining genera (t -tests with log-transformed data). Similar patterns were evident when all genera that increased and declined in density (not just those that changed significantly) were compared. Collectively, these trends suggest that genera with higher absolute growth rates, including many canopy and emergent trees but not pioneers, are increasing at the expense of slower-growing genera, which include many smaller, old-growth subcanopy trees.

The tree community is also changing in taxonomic composition. The increasing genera are dominated (57%) by three families (Leguminosae, Lecythidaceae, Sapotaceae) that are not represented

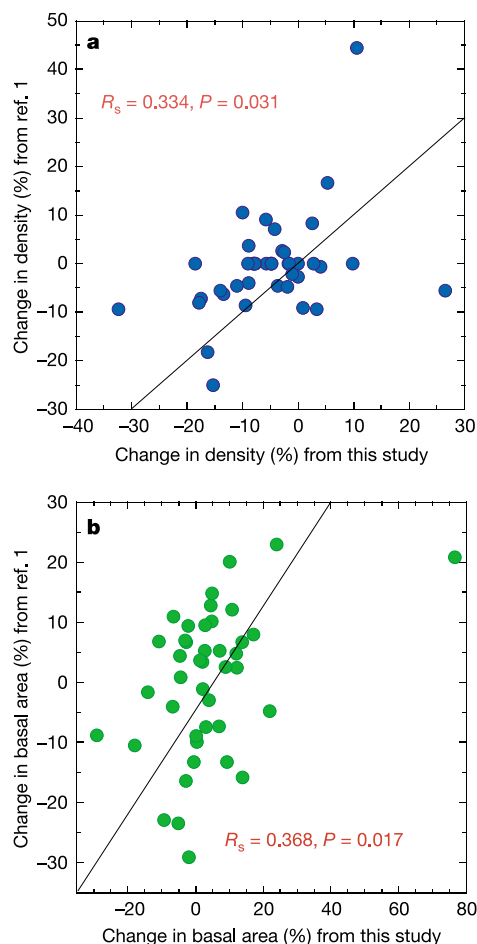


Figure 1 Mean percentage changes in (a) population density and (b) basal area of 42 Amazonian tree genera in two different long-term studies. (Correlation coefficients are for Spearman rank tests.) Data are from the 18 one-hectare plots in this study and the three one-hectare plots of ref. 1. The diagonal line in each figure shows $y = x$.

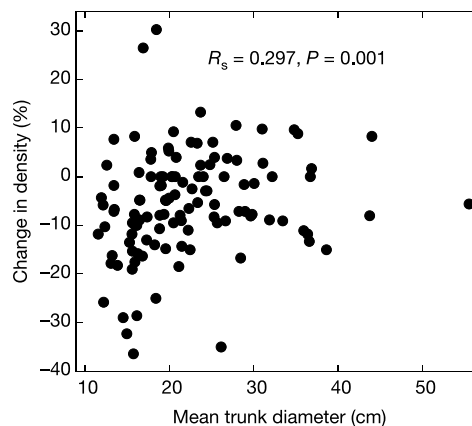


Figure 2 Relationship between tree size and long-term population change for Amazonian tree genera. (The correlation coefficient is for a Spearman rank test.)

among declining genera, whereas most (64%) declining genera are in families (Arecaceae, Annonaceae, Chrysobalanaceae, Moraceae, Myristicaceae) that are poorly represented among increasing genera (Table 1).

To help identify the underlying causes of these alterations, we assessed dynamical changes in the tree communities. We divided census data for each plot into two roughly equal intervals (1984–91 and 1992–99) and then contrasted overall rates of tree mortality and recruitment between the two intervals. Both rates rose markedly from interval 1 to interval 2 (Fig. 3); thus our forests clearly became more dynamic over time. Mortality and recruitment rates did not rise significantly for the increasing and declining genera, although the latter consistently had higher mortality than recruitment (Fig. 3).

Moreover, for 87% of genera, rates of trunk growth accelerated between intervals 1 and 2 (Fig. 4). This is the first demonstration of enhanced growth across a wide range of tropical tree genera, and is consistent with stand-level increases in tree growth across South American forests¹¹. Notably, the average increase in absolute growth rate was higher among increasing genera than declining genera (0.55 ± 0.49 versus 0.19 ± 0.17 mm yr⁻¹); the average for genera showing no significant change was intermediate (0.41 ± 0.50 mm yr⁻¹). This difference did not occur solely because increasing genera were often large in size and decreasing genera often small: in relative terms, growth accelerated much more in increasing (57%) than decreasing (22%) genera.

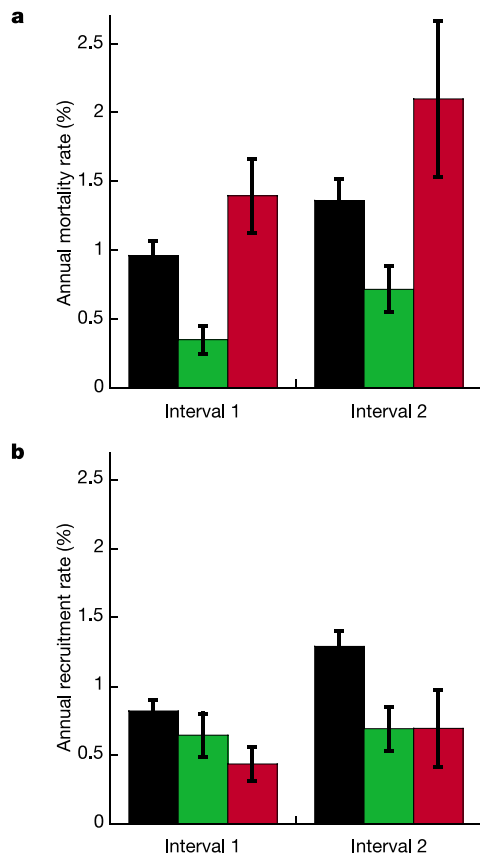


Figure 3 Mortality (a) and recruitment (b) rates (± 1 s.e.m.) for increasing and declining genera. Data are shown for all trees (black bars), for 13 genera that increased in basal area (green bars), and for 13 genera that declined in density (red bars). Overall mortality ($t = -2.38$, d.f. = 17, $P = 0.03$) and recruitment ($t = -4.45$, d.f. = 17, $P = 0.0003$) accelerated from interval 1 (around 1984–91) to interval 2 (around 1992–99). However, there was no significant change over time ($P > 0.11$) in mortality or recruitment for the increasing and declining genera (paired t -tests).

At least four different phenomena might account for these forest-wide changes in composition and dynamics. First, the forests of our study area might be in a state of disequilibrium because of ongoing recovery from past disturbance, leading to shifts over time in tree composition. However, the only natural disturbance likely to operate over such large (300 km²) spatial scales would be a major forest fire, and charcoal and phytolith data suggest that our study area has been continuously forested for at least 4,500 yr²³. Other forms of disturbance, such as windbursts from convectional thunderstorms and wet-season flooding, are patchy and localized, and thus are unlikely to account for the pervasive changes we detected.

Second, the observed changes might reflect differential vulnerability of trees to El Niño-related droughts^{15,16}, which increased in frequency during the last century²⁴, possibly because of global warming²⁵. We found little direct support for this proposition in two separate analyses (Supplementary Tables 3 and 4). The drought hypothesis, however, requires further examination, because strong droughts have apparently caused shifts in tree-community composition in Panama¹⁵.

Third, the forests might be responding to multi-decadal changes in rainfall that affect forest productivity and species composition. Drier conditions in rainforests may increase tree growth and reproduction^{26,27}, possibly because cloud cover is reduced, increasing available sunlight for light-limited trees. Nonetheless, no significant trends in central-Amazon rainfall were evident during our study (Supplementary Table 5) or in the preceding decades²⁸, providing little support for this hypothesis.

Finally, changes in floristic composition could be driven by accelerated forest productivity. We believe the most likely cause of higher productivity is rising atmospheric CO₂ levels^{2–13}, although other agents (such as possible increases in solar radiation from reduced tropical cloudiness^{17,29}, and higher airborne nutrient deposition¹⁸ from increasing forest fires) are also plausible. Of all the hypothesized factors, accelerated productivity best explains key observations of this study: (1) that tree growth, mortality and recruitment have increased markedly, all of which can result from higher productivity^{2–13}; (2) that many faster-growing genera are increasing in basal area, possibly because fast-growing trees show stronger growth enhancement under increased CO₂ (refs 2, 3, 13); and (3) that forests are experiencing nonrandom changes in

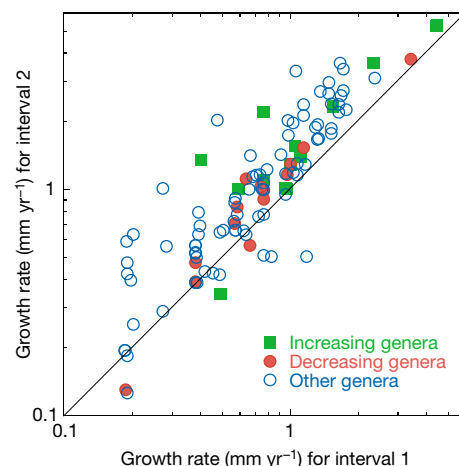


Figure 4 Comparison of median growth rates of Amazonian tree genera between interval 1 (around 1984–91) and interval 2 (around 1992–99). The diagonal line shows $y = x$. Growth rates accelerated markedly over time for all genera ($t = -9.74$, d.f. = 114, $P < 0.00001$; paired t -test), and accelerated significantly more for increasing than for decreasing genera ($t = 2.45$, d.f. = 24, $P = 0.022$; t -test for unequal variances).

species composition, with fast-growing canopy and emergent genera evidently gaining a competitive advantage over smaller, slower-growing genera. That rapidly growing pioneers have not increased in abundance is surprising, but these species usually establish themselves in large treefall gaps^{2,3,13}, which may be uncommon in our study area because mortality is greatest among small trees. The group most likely to decline further, we suggest, is old-growth subcanopy species, a highly diverse assemblage that are notable for their slow growth, dense wood and ability to reproduce in full shade³⁰.

These changes could have both local and global consequences. Undisturbed Amazonian forests appear to be functioning as an important carbon sink^{7–11}, helping to slow global warming, but pervasive changes in tree communities could modify this effect. In particular, increases in forest carbon storage may be slowed by the tendency of canopy and emergent trees to produce wood of reduced density as their size and growth rate increases³⁰, and by the decline of densely wooded subcanopy species. Forest-wide alterations of tree communities, which sustain assemblages of often-specialized pollinators, herbivores, symbiotic fungi and other species, could also have serious ecological repercussions for the diverse Amazonian biota. Further studies are urgently needed to determine whether comparably large shifts in tree communities are occurring throughout the tropics, and to identify the environmental agents driving these changes. □

Methods

Changes in tree density and basal area

Our null hypothesis was that each tree genus exhibited no significant change in population density or basal area, which is appropriate because total density and total basal area of trees did not change significantly during our study. For each genus, change in mean population density (λ) was defined as $\frac{N_T - N_0}{N_0}$, where N_T is the final population density and N_0 is the initial density (N for each genus was found by averaging population density across all 18 plots). To estimate confidence limits for λ , we bootstrapped across plots. Eighteen plots were drawn at random, with replacement, and λ calculated each time (the same set of plots was used to find N_T and N_0). From 1,000 replicates, the 5th and 95th ranking values of λ were taken as the 99% confidence limits. Observed values of λ that fell outside this range were considered significant at the $P \leq 0.01$ level (using a two-tailed test). Because this method is unreliable for genera occurring in a small number of plots, we restricted analyses to genera present in eight or more plots during our initial census (at this frequency, all genera exhibited reasonably stable estimates for recruitment, mortality and growth). We used the same method to test for changes in basal area of each species.

Correlated changes among plots

We used randomization tests to determine whether the observed changes in density and basal area for all 115 tree genera were more similar among our 18 plots than was expected by chance. To do this we selected nine plots at random and determined the mean per cent change in density for each genus in the plots, and then compared these values to the mean per cent change for each genus in the other nine plots, using Pearson correlations. We repeated this 1,000 times, using random combinations of plots each time. The mean and standard error of the mean (s.e.m.) for the 1,000 correlations were determined, which were then used to calculate a Z statistic ($Z = \text{mean} \times \text{s.e.m.}^{-1}$). We used a one-tailed Z-test to determine whether the mean value of the observed correlations was significantly greater than zero. The same procedure was used to test for changes in basal area.

Independent study

The independent study¹ used methods that were nearly identical to ours. Three one-hectare plots were censused in August 1991 and again in September 1999. A total of 2,085 trees were recorded in the plots, of which 97.1% were identified to genus. To minimize effects of small sample sizes, we included in the analysis the 42 genera with at least ten individuals in the three plots (Fig. 1). All of these genera were present in at least half of our 18 plots.

Ecological traits of tree genera

For most of the 115 genera in this study, data on growth form and successional status were gleaned from a variety of published and online data sources as well as personal knowledge of the authors (Supplementary Tables 1 and 2). Data on other ecological traits (for example, seed-dispersal syndrome, fruit type) were not available for our genera on a consistent basis. Estimates for median and maximum growth rates, mortality and recruitment rates, and mean trunk diameter were derived from demographic data from our long-term study (Supplementary Table 1). Distributional data on locally occurring species within each significantly changing genus, with respect to major rainfall zones in Amazonia, were mostly derived from online sources (Supplementary Table 3). Finally, an index of drought tolerance for 30 abundant tree species was derived from published and unpublished data from our 18 plots and from other nearby plots in the same study area (Supplementary Table 4).

Changes in forest dynamics and growth

Stand-level rates of annual mortality and recruitment, and the annual rate of trunk growth for individual tree genera were generated for two largely non-overlapping intervals (approximately 1984 to 1991 and 1992 to 1999). For our 18 plots, the first interval averaged 7.6 ± 2.5 yr in duration, whereas the second interval averaged 7.4 ± 0.9 yr in duration; the first interval was more variable in length because the plots were initially established over a six-year period, from 1981 to 1987. Annualized mortality and recruitment data for each plot and interval were estimated using logarithmic models.

To calculate annual growth rates for each genus, the mean annual growth of each tree was determined by subtracting its initial DBH from its final DBH, and dividing by the number of years. The median growth rate was then determined for all trees within the genus. Maximum growth rate was also calculated for each genus, and to reduce the effects of outliers the upper decile of the rates was used as an estimate of maximum growth rate. Growth rates were calculated only for genera that had at least ten live stems in each of the first and second intervals.

Received 22 September 2003; accepted 2 February 2004; doi:10.1038/nature02383.

- de Oliveira, A. A. & Mori, S. A. A central Amazonian terra firma forest. I. High tree species richness on poor soils. *Biodivers. Conserv.* **8**, 1219–1244 (1999).
- Reekie, E. G. & Bazzaz, F. A. Competition and patterns of resource use among seedlings of five tropical trees grown at ambient and elevated CO_2 . *Oecologia* **79**, 212–222 (1989).
- Winter, K. & Lovelock, C. E. Growth responses of seedlings of early and late successional tropical forest trees to elevated atmospheric CO_2 . *Flora* **194**, 221–227 (1999).
- Phillips, O. L. & Gentry, A. H. Increasing turnover through time in tropical forests. *Science* **261**, 954–958 (1994).
- Phillips, O. L. *et al.* Pattern and process in Amazon tree turnover, 1976–2001. *Phil. Trans. R. Soc. B.* (in the press).
- Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Grace, J. *et al.* Carbon dioxide uptake by an undisturbed tropical rain forest in southwest Amazonia, 1992 to 1993. *Science* **270**, 778–780 (1994).
- Phillips, O. L. *et al.* Changes in the carbon balance of tropical forest: Evidence from long-term plots. *Science* **282**, 439–442 (1998).
- Baker, T. R. *et al.* Increasing biomass in Amazonian forest plots. *Phil. Trans. R. Soc. B.* (in the press).
- Malhi, Y. *et al.* Carbon dioxide transfer over a central Amazonian rain forest. *J. Geophys. Res.* **103**, 31593–31612 (1998).
- Lewis, S. L. *et al.* Concerted changes in tropical forest structure and dynamics: Evidence from 50 South American long-term plots. *Phil. Trans. R. Soc. B.* (in the press).
- Phillips, O. L. *et al.* Increasing dominance of large lianas in Amazonian forests. *Nature* **418**, 770–774 (2002).
- Körner, C. Tropical forests in a CO_2 -rich world. *Clim. Change* **39**, 297–315 (1998).
- Clark, D. A., Piper, S. C., Keeling, C. D. & Clark, D. H. Tropical rain forest tree growth and atmospheric carbon dynamics linked to interannual temperature variation during 1984–2000. *Proc. Natl Acad. Sci. USA* **100**, 5852–5857 (2003).
- Condit, R., Hubbell, S. P. & Foster, R. B. Assessing the response of plant functional types to climatic change in tropical forests. *J. Veg. Sci.* **7**, 405–416 (1996).
- Tian, H. *et al.* Effect of interannual climate variability on carbon storage in Amazonian ecosystems. *Nature* **396**, 664–667 (1998).
- Wielicki, B. A. *et al.* Evidence for large decadal variability in tropical mean radiative energy budget. *Science* **295**, 841–844 (2002).
- Artaxo, P. *et al.* Dry and wet deposition in Amazonia: from natural biogenic aerosols to biomass burning impacts. *Int. Glob. Atmos. Chem. News.* **27**, 12–16 (2003).
- Laurance, W. F. *et al.* Biomass collapse in Amazonian forest fragments. *Science* **278**, 1117–1118 (1997).
- Laurance, W. F. *et al.* Rainforest fragmentation kills big trees. *Nature* **404**, 836 (2000).
- ter Steege, H. & Hammond, D. S. Character convergence, diversity, and disturbance in tropical rain forest in Guyana. *Ecology* **82**, 3197–3212 (2001).
- Casper, B. B., Heard, S. B. & Apanius, V. Ecological correlates of single-seededness in a woody tropical flora. *Oecologia* **90**, 212–217 (1992).
- Piperno, D. R. & Becker, P. Vegetation history of a site in the central Amazon Basin derived from phytolith and charcoal records from natural soils. *Quat. Res.* **45**, 202–209 (1996).
- Dunbar, R. B. El Niño—clues from corals. *Nature* **407**, 956–959 (2000).
- Timmermann, A. *et al.* Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature* **398**, 694–697 (1999).
- Wright, S. J. *et al.* The El Niño Southern Oscillation, variable fruit production, and famine in a tropical forest. *Ecology* **80**, 1632–1642 (1999).
- Clark, D. A. & Clark, D. B. Climate-induced annual variation in canopy tree growth in a Costa Rican tropical rain forest. *J. Ecol.* **82**, 865–872 (1994).
- Malhi, Y. & Wright, J. Spatial patterns and recent trends in the climate of tropical forest regions. *Phil. Trans. R. Soc. Lond. B.* (in the press).
- Lewis, S. L., Malhi, Y. & Phillips, O. L. Fingerprinting the impacts of global change on tropical forests. *Phil. Trans. R. Soc. Lond. B.* (in the press).
- Thomas, S. C. Asymptotic height as a predictor of growth and allometric characteristics in Malaysian rain forest trees. *Am. J. Bot.* **83**, 556–566 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the many taxonomic specialists and field technicians who collected and identified plant material during the study, and the A. W. Mellon Foundation, the NASA-LBA Program, the Conservation, Food and Health Foundation, the US National Science Foundation, the MacArthur Foundation, the National Institute for Amazonian Research (INPA), and the Smithsonian Institution for support. Comments from O. L. Phillips, J. A. Pounds, S. L. Lewis, S. J. Wright, K. Winter, T. R. Baker, R. K. Didham, R. Harrison, T. A. Kursar and P. D. Coley improved the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.F.L. (laurancew@tivolisi.edu).

Sustainable trophy hunting of African lions

Karyl Whitman, Anthony M. Starfield, Henley S. Quadling & Craig Packer

Department of Ecology, Evolution and Behavior, University of Minnesota, 1987 Upper Buford Circle, Saint Paul, Minnesota 55108, USA

In most species, sport hunting of male trophy animals can only reduce overall population size when the rate of removal of males is so high that females can no longer be impregnated¹. However, where males provide extensive paternal care, the removal of even a few individuals could harm the population as a whole^{2,3}. In species such as lions, excessive trophy hunting could theoretically cause male replacements (and associated infanticide^{4,5}) to become sufficiently common to prevent cubs reaching adulthood. Here we simulate the population consequences of lion trophy hunting using a spatially explicit, individual-based, stochastic model parameterized with 40 years of demographic data from northern Tanzania. Although our simulations confirm that infanticide increases the risk of population extinction, trophy hunting could be sustained simply by hunting males above a minimum age threshold, and this strategy maximizes both the

quantity and the quality of the long-term kill. We present a simple non-invasive technique for estimating lion age in populations lacking long-term records, and suggest that quotas would be unnecessary in any male-only trophy species where age determination could be reliably implemented.

Male lions reach sexual maturity at about 2.5 yr of age and live to a maximum of about 15 yr in nature⁶. The lion's mane reaches full size at about 4 yr (ref. 7), and peak reproductive success is attained by about 8 yr (ref. 8). African lions live in stable social groups ('prides') containing an average of six breeding females and a coalition of 2–3 adult males. The resident coalition sires all cubs born during their tenure⁹, but most coalitions only remain resident for about 2 yr on average—long enough to rear a single cohort of young to independence¹⁰. Rather than wait for mothers with dependent offspring to rear their current brood, incoming males typically kill all cubs ≤ 9 months of age and evict older subadults when they first take over a pride^{4,5}. Trophy hunting is expected to increase the rate of male takeovers, as larger coalitions dominate smaller ones¹¹ and the loss of even one male from a resident coalition renders it more vulnerable to being ousted¹².

We developed a comprehensive simulation model that tracks the fate of each individual in a population¹³ (see Methods), and we present results based on 'populations' comprising a maximum of ten prides of ≤ 9 females per pride. Outcomes of hunting should be most sensitive to factors that limit population size: the maximum number of prides in the population, maximum pride size, and the incidence of infanticide. We therefore ran simulations of populations containing a maximum of five prides with ≤ 10 females and ten prides with ≤ 7 females, and our conclusions were unchanged. The impact of infanticide is emphasized below (see Fig. 1). At each six-month time step, animals survive and breed according to probabilities observed in the long-term lion studies in the Serengeti National Park and in Ngorongoro Crater^{14,15}. Demographic parameters depend on the age, sex and social status of individual lions. Probability of female recruitment depends on the number of adult females in the pride, whereas probability of male takeovers depends on the size and age of resident coalitions versus challenging nomadic coalitions. An emerging property of these interactions is a density-dependence that leads to a quasi-equilibrium where the total population size fluctuates slightly (with demographic stochasticity) around a well-defined average. Preliminary trials started the populations with an arbitrary set of individuals and an arbitrary age

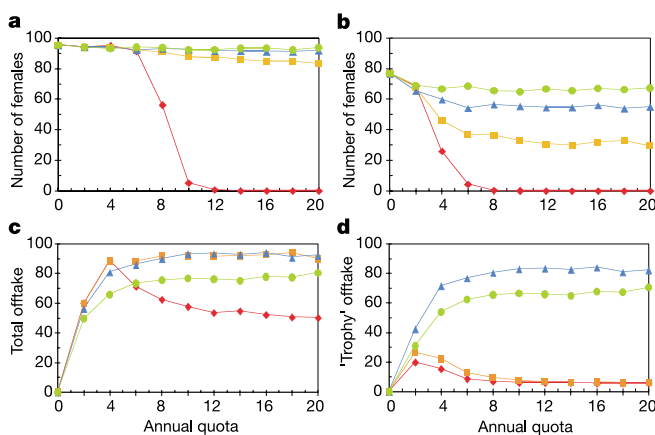


Figure 1 Effects of trophy hunting as a function of quota size and male age. Average outcome after 100 runs is shown for shooting males of the following ages: ≥ 3 yr old (red), ≥ 4 yr (orange), ≥ 5 yr (blue), ≥ 6 yr (green). **a**, Number of adult females after 30 yr in hypothetical populations where males are non-infanticidal. **b**, Number of females in infanticidal populations; note that infanticidal populations are smaller and more vulnerable to trophy hunting. **c**, Total number of males harvested over 30 yr in infanticidal populations. **d**, Total number of 5–6-yr-old 'trophies' harvested in infanticidal populations.

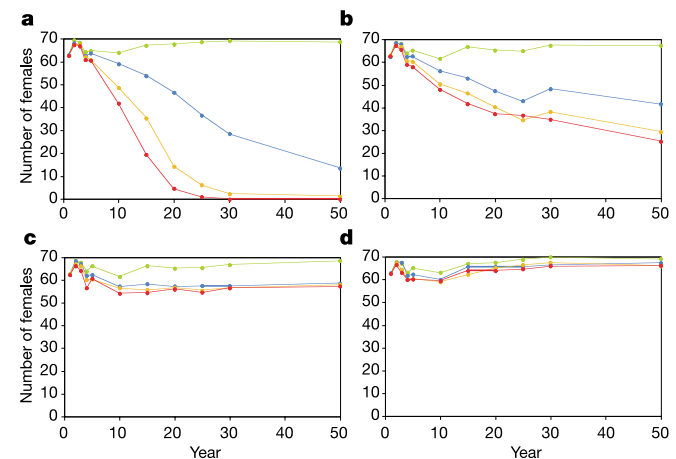


Figure 2 Female population size through time as a function of quota size and male age in infanticidal populations. Red indicates average outcome over 100 runs from an annual quota of 10 males, orange from a quota of 6 males, blue from a quota of 4, and green from a quota of 2. **a–d**, Female population size when hunters shoot males that are ≥ 3 yr old (**a**), ≥ 4 yr (**b**), ≥ 5 yr (**c**) and ≥ 6 yr (**d**).

distribution, and proceeded (without trophy hunting) until the population reached a stable size and age-structure that served as the standardized starting point for all simulations.

Trophy-hunting strategies were varied in two ways. First, the minimum age of eligible males ranged from ≥ 3 yr to ≥ 6 yr. Second, quotas could range from 0 to 20 males per year. Individual males were removed according to the age minimum, but at random with respect to pride or nomadic status. Figure 1 illustrates the effects of these strategies after 30 yr (as start-up populations required 20–25 yr to equilibrate). Figure 1a and b shows that if off-take included males as young as three years of age, the female population would invariably collapse if quotas were too high. But as the minimum age of trophy males was raised, the chances of population persistence increased markedly—to the point where removing males ≥ 6 yr in age had no substantial effect, regardless of quota size. Note that the sensitivity to over-hunting is exacerbated by infanticide: if we eliminate infanticide from the model, the resultant populations are always larger regardless of hunting strategy (Fig. 1a)—but removing too many young males will still drive the population to extinction because females are eventually unable to mate. Figure 2 shows the average size of the female population through time as a function of male age and quota size. Shooting too many young males leads to an inevitable decline in population size, whereas quota size is irrelevant when hunting is restricted to older males. Running the simulations for 50 yr in infanticidal populations, quotas of more than two ≥ 3 - or ≥ 4 -yr-old males per year produced at least one case of extinction per 100 runs, whereas there were no extinctions when hunting was restricted to ≥ 5 - or ≥ 6 -yr-olds.

By only removing ≥ 5 - or ≥ 6 -yr-old males, younger males have the opportunity to remain resident long enough to rear a cohort of young. By removing males in a manner that sustains population growth, the population can yield more males in the long run. Figure 1c and d shows that the cumulative number of males harvested over 30 yr is also sensitive to male age and quota, but, again, the detrimental impact of trophy hunting is largely avoided by maintaining a minimum age of 5–6 yr, and the total number of high-quality ‘trophies’ (large-maned males of 5–6 yr) taken at random from these populations is also highest. Hunting fees are highest in reserves with greater opportunities to shoot trophy animals, so it is noteworthy that the most impressive males are so ‘expendable’ to the population. Figure 3 shows the expected

annual harvest of ‘trophy males’ through time. By restricting off-take to ≥ 5 - or ≥ 6 -yr-old males, hunters harvest a greater cumulative number of trophy males and are likely to gain a steadier off-take each year.

Hunters often estimate male age on the basis of mane size or colouration, but these phenotypes are only loosely correlated with age and vary greatly across the geographic range of the lion⁷. The most reliable index in the Serengeti/Ngorongoro lions is the extent of dark pigmentation in the tip of the nose, which becomes increasingly freckled with age (Fig. 4). Individual variation in nose colouration is sufficiently low that age can be estimated up to 8–9 yr (Table 1). The noses of 5-yr-old males are 50% black, so a simple rule of thumb would be to restrict all trophy hunting to males with noses that are more than half black. However, the noses of Ngorongoro males darkened more slowly than Ngorongoro females and Serengeti males/females (Fig. 4), thus site-specific data may be necessary to provide accurate age estimates. However, in areas with slower rates of nose darkening, the 50% rule would safely restrict hunting to ≥ 8 -yr-old males.

Detailed predictions of these simulations should be interpreted

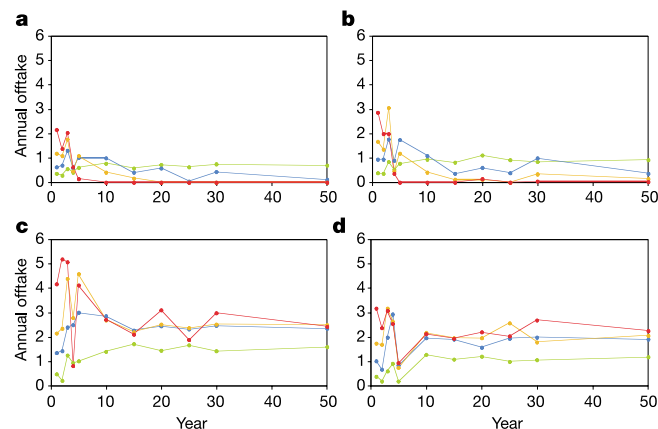


Figure 3 Annual off-take of 5–6-yr-old ‘trophy’ males as a function of quota size and male age in infanticidal populations. Red indicates average outcome of 100 runs from an annual quota of 10 males, orange from a quota of 6 males, blue from a quota of 4, and green from a quota of 2. Annual off-take of trophies when hunters shoot males that are ≥ 3 yr old (a), ≥ 4 yr (b), ≥ 5 yr (c) and ≥ 6 yr (d).

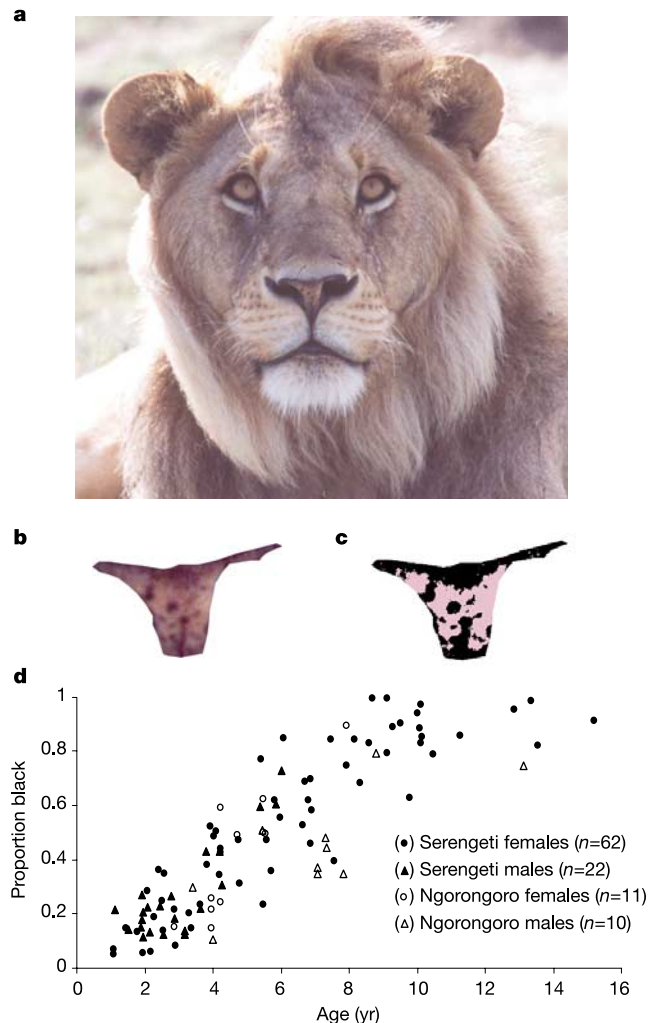


Figure 4 Age estimation for adult lions using nose colouration. a, Identification photograph of a 3-yr-old Serengeti male. b, Excised photo of nose tip. c, GIS rendering of nose colouration. d, Age-change of nose colouration for males and females in two separate populations. After controlling for age, there was no effect of sex on nose colour in the Serengeti, but Ngorongoro males had lighter noses than Ngorongoro females ($P = 0.0485$) and Serengeti males ($P = 0.0281$).

with caution, because they rely on demographic patterns in two adjacent populations in northern Tanzania and may not apply to every situation. Nevertheless, our results do point to possible conflicts between conservation and consumption. Figures 1b and 2 suggest that lion populations will be larger if trophy hunting is restricted to males ≥ 6 yr, but this strategy entails an opportunity cost: long-term harvests are highest from hunting males with ages ≥ 5 yr (Fig. 1c and d, Fig. 3). Our analysis also treats the lion population as a 'closed system', sustained solely by birth rates of the resident females, which exemplifies populations most at risk from overexploitation. 'Open' systems involve game reserves immediately adjacent to national parks where excess males and females can freely disperse from protected areas into low density hunting areas. Our analysis also assumes that trophies are removed at random according to residence status. If hunters followed a more sophisticated approach by selectively targeting nomadic males or resident males whose cubs have recently reached independence, trophy hunting could greatly reduce the incidence of infanticide, and population sizes would approach the levels predicted by Fig. 1a.

Assuming an average pride range of 100 km², our simulated population of ten prides would yield fewer than three males per 1,000 km² per year (Fig. 1c). Hunting quotas in 34 reserves in Tanzania in 1995 were set at an average of 3.8 males per 1,000 km² (median, 2.8; range, 0.5–13.4). Although the expectations from these concessions were generally set at an appropriate order of magnitude, quotas have increased in many Tanzanian reserves since 1995, and quotas are even higher in parts of Africa where lion population densities are considerably lower. Therefore, realistic lion harvests for most areas may not generate sufficient revenue to be economically viable in the long term unless concessions establish an auction system¹⁶ or sell hunting 'opportunities' (which may or may not succeed in killing a lion) rather than attempt to base their income exclusively on successful hunts.

Our model shows clearly that age is a critical variable enabling the persistence of trophy species with extensive paternal care. As long as hunting is restricted to a safe minimum age (and the rule is honestly enforced¹⁷), there is no risk of setting excessive quotas even in areas where it is impossible to estimate the overall population size¹⁸. Lions' noses continue to darken until they are at least 9 yr old, so rough estimates of male age can easily be made (Table 1). Thus despite the complexity of lion social structure and the complex consequences of male removal, populations can be sustained by following a simple harvest rule, combined with a simple technique for age-assessment.

Although our model was designed to examine the impact of trophy hunting in only one species, our findings have broad significance. First, highly valued 'trophy' animals are individuals with unusually well-developed secondary sexual characteristics, such as large antlers, horns, tusks or mane. Although these traits

generally grow with age, high-quality individuals may show precocious physical development and thus be shot before they are able to breed—with negative evolutionary effects on the population as a whole¹⁹. Thus it is particularly important to assess age independently from the 'trophy' phenotype, and to set age thresholds high enough to ensure successful breeding by the best males. Second, the hunting industry has always been based on some sort of quota system, even though quotas are generally viewed as arbitrary and difficult to enforce^{16,20,21}. Our analysis suggests that quotas could eventually become irrelevant to the conservation of lions hunting is restricted to males. With basic information on breeding biology and social behaviour (such as infanticide and other quirks of male behaviour²²), an age-threshold criterion could be calculated that would minimize the adverse effects from killing all of the eligible males each year. By excluding the younger adult males from the annual kill, females would always be impregnated and vulnerable offspring would generally be protected by their fathers. □

Methods

Simulation model

In the Windows-driven C++ model¹², female lions and their dependent offspring are organized into 'prides' that defend spatially arranged and interconnected territories. The model ignores environmental stochasticity, so the maximum number of territories and maximum pride size is held constant for a given set of simulations. The model distinguishes between sex and age class (cubs <6 months, cubs 6–12 months, cubs 12–24 months, subadult males and females, adult males and females) and tracks individuals by social and reproductive status. Only 3–13-yr-old females produce cubs; females are not able to breed again until they lose their entire litter or their surviving offspring reach 2 yr of age. Males are classified as either subadult, nomadic or resident; lone males may join up with other lone males or groups of two. Nomadic and subadult males move freely between and within pride territories a specified number of times per time step, but do not breed with females. Residents may be affiliated with up to three prides at once. Competition for pride residence is determined by using a competition matrix²³ that weights overall competitive strength according to male age and coalition size. Cubs are killed with an age-specific probability when new males first enter a pride.

At each time step, the model simulates cub production using a random number for each eligible female that draws her litter size from a distribution, determines individual survival (assigned by random number according to the observed survival rate for animals of that age, sex and social status—see Supplementary Information), updates ages for survivors, organizes 2-yr-old males in each pride into subadult male groups, promotes 3-yr-old males into nomadic groups, and determines the fate of subadult females. Recruitment of females into their natal pride depends on the number of adult females already in the pride and the specified upper limit for that pride (which can be temporarily exceeded by no more than two females). If the subadult females cannot be accommodated in their natal pride, they are allowed to search for empty territories, but they die if they cannot find any vacancies. At 'equilibrium', the simulated populations closely matched real populations in terms of the male:female:cub ratio, the average age of resident males, and the size range of subadult cohorts. To test the effects of trophy hunting, each simulation ran for 30 or 50 yr with 100 replicates. The initial age structure, reproductive history, pride affiliation, and distribution were identical for each replicate. Harvesting occurred at random (within the rules) every 6 months, although it was not always possible to meet the quota.

Nose assessments

Close-up colour photographs ($n = 189$) were taken of 105 known-aged lions, including 73 females and 32 males from the Serengeti National Park and Ngorongoro Crater, Tanzania, between 1999 and 2002. Each photograph was first digitized at high resolution into a .tif file, and the fleshy part of the nose ('nose tip') from each image was excised using Adobe Photoshop 4.01 LE. We then used the Spatial Analyst extension of ESRI Arcview 3.2 to rasterize each cut-out nose tip and assign each newly created 'grid' a range of colour values. By limiting the colour values to either 'black' or 'not black', the nasal pigmentation pattern was 'mapped' and quantified for the percentage of readable pixels that contained 'black'. We used linear multiple regression to assess the effect of age, sex, habitat and population on pigmentation. In cases where multiple photographs were available for the same individual, the mean pigmentation score was regressed against mean age.

Received 25 November 2003; accepted 4 February 2004; doi:10.1038/nature02395.

Published online 22 February 2004.

Table 1 Statistical relationship between nose blackness and age

Proportion black	Estimated age in years (s.e.)	95% p.i.	75% p.i.	50% p.i.
0.10	2.66 (1.24)	0.17–5.15	1.21–4.10	1.81–3.50
0.20	3.25 (1.24)	0.77–5.72	1.81–4.69	2.41–4.09
0.30	3.84 (1.23)	1.37–6.30	2.40–5.27	3.00–4.68
0.40	4.42 (1.23)	1.97–6.89	3.00–5.86	3.59–5.26
0.50	5.02 (1.23)	2.56–7.48	3.59–6.45	4.18–5.85
0.60	5.61 (1.23)	3.14–8.07	4.18–7.04	4.77–6.44
0.70	6.20 (1.23)	3.73–8.66	4.77–7.63	5.36–7.04
0.80	6.79 (1.24)	4.32–9.26	5.35–8.23	5.95–7.63
0.90	7.38 (1.24)	4.90–9.87	5.94–8.82	6.54–8.22
1.00	7.97 (1.25)	5.58–10.47	6.52–9.42	7.12–8.82

s.e., standard error; p.i., predicted interval. Predicted values are based upon the least-squares regression of a truncated data set for 63 known-aged females in the Serengeti and Ngorongoro aged ≤ 10 yr ($y = 2.0667 + 5.9037 \arcsin(x)$; $r^2 = 0.75$, $P < 0.0001$). Predicted intervals at 95%, 75% and 50% are included for upper and lower bounds.

1. Milner-Gulland, E. J. *et al.* Reproductive collapse in saiga antelope harems. *Nature* **422**, 135 (2003).
2. Swenson, J. E. *et al.* Infanticide caused by hunting of male bears. *Nature* **386**, 450–451 (1997).
3. Greene, C., Umbanhowar, J., Mangel, M. & Caro, T. in *Behavioral Ecology and Conservation Biology* (ed. Caro, T.) 271–305 (Oxford Univ. Press, Oxford, 1998).
4. Bertram, B. C. R. Social factors influencing reproduction in wild lions. *J. Zool.* **177**, 463–482 (1975).
5. Packer, C. Infanticide is no fantasy. *Am. Anthropol.* **102**, 829–831 (2000).
6. Schaller, G. B. *The Serengeti Lion* (Univ. Chicago Press, Chicago, 1972).

7. West, P. M. & Packer, C. Sexual selection, temperature and the lion's mane. *Science* **297**, 1339–1343 (2002).
8. Packer, C. *et al.* in *Reproductive Success* (ed. Clutton-Brock, T. H.) 363–383 (Univ. Chicago Press, Chicago, 1988).
9. Packer, C., Gilbert, D., Pusey, A. E. & O'Brien, S. J. A molecular genetic analysis of kinship and cooperation in African lions. *Nature* **351**, 562–565 (1991).
10. Packer, C. & Pusey, A. E. Male takeovers and female reproductive parameters: a simulation of oestrus synchrony in lions (*Panthera leo*). *Anim. Behav.* **31**, 334–340 (1983).
11. Bygott, J. D., Bertram, B. R. C. & Hanby, J. P. Male lions in large coalitions gain reproductive advantages. *Nature* **282**, 838–840 (1979).
12. Grinnell, J., Packer, C. & Pusey, A. E. Cooperation in male lions: kinship, reciprocity or mutualism? *Anim. Behav.* **49**, 95–105 (1995).
13. Quadling, H. S. & Starfield, A. M. Exploiting object-orientated programming structures in the quest for an individual-based lion population model with an attractive user interface. *S. Afr. J. Sci.* **98**, 449–454 (2002).
14. Packer, C., Tatar, M. & Collins, D. A. Reproductive cessation in female mammals. *Nature* **392**, 807–811 (1998).
15. Packer, C., Pusey, A. E. & Eberly, L. Egalitarianism in female African lions. *Science* **293**, 690–693 (2001).
16. Leader-Williams, N., Kayera, J. A. & Overton, G. L. *Tourist Hunting in Tanzania* (IUCN Species Survival publication no. 14, Gland, Switzerland, 1996).
17. Smith, R. J., Muir, R. D. J., Walpole, M. J., Balmford, A. & Leader-Williams, N. Governance and the loss of biodiversity. *Nature* **426**, 67–70 (2003).
18. Hillborn, R. & Walters, C. J. *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty* (Chapman & Hall, New York, 1992).
19. Coltman, D. W. *et al.* Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655–658 (2003).
20. Caro, T. M. *et al.* The impact of tourist hunting on large mammals in Tanzania: an initial assessment. *Afr. J. Ecol.* **36**, 321–329 (1998).
21. Sutherland, W. J. Sustainable exploitation: a review of principles and methods. *Wildl. Biol.* **7**, 131–140 (2001).
22. Slotow, R., Van Dyk, G., Poole, J., Page, B. & Klocke, A. Older bull elephants control young males. *Nature* **408**, 425–426 (2000).
23. Starfield, A. M., Furniss, P. R. & Smuts, G. L. in *Dynamics of Large Mammal Populations* (ed. Fowler, C. W.) 121–134 (Wiley, New York, 1981).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the government of Tanzania for permission to conduct research; K.W. was supported by the Wildlife Conservation Society, Big Game Special Projects Foundation, John D. & Catherine T. MacArthur Foundation, Graduate School of the University of Minnesota, Dayton-Wilkie Foundation, and Global Wildlife Trust; C.P. was supported by the McKnight Foundation and NSF LTREB and Animal Behavior programmes. We thank T. Gelatt, P. West, G. Hopcraft, M. Craft, B. Kissui, L. Frank, S. Mduma, T. Ramme, W. Testa, T. Coulson and N. Leader-Williams for advice, assistance and discussion.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.P. (packer@cbs.umn.edu).

Fitness benefits of prolonged post-reproductive lifespan in women

Mirkka Lahdenperä¹, Virpi Lummaa^{2,3}, Samuli Helle¹, Marc Tremblay⁴ & Andrew F. Russell²

¹Section of Ecology, Department of Biology, University of Turku, FIN-20014 Turku, Finland

²Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

³Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

⁴Interdisciplinary Research Group on Demography and Genetic Epidemiology, University of Quebec, Chicoutimi, Quebec G7H 2B1, Canada

Most animals reproduce until they die, but in humans, females can survive long after ceasing reproduction^{1,2}. In theory, a prolonged post-reproductive lifespan will evolve when females can gain greater fitness by increasing the success of their offspring than by continuing to breed themselves^{3–6}. Although

reproductive success is known to decline in old age^{1–6}, it is unknown whether women gain fitness by prolonging lifespan post-reproduction. Using complete multi-generational demographic records, we show that women with a prolonged post-reproductive lifespan have more grandchildren, and hence greater fitness, in pre-modern populations of both Finns and Canadians. This fitness benefit arises because post-reproductive mothers enhance the lifetime reproductive success of their offspring by allowing them to breed earlier, more frequently and more successfully. Finally, the fitness benefits of prolonged lifespan diminish as the reproductive output of offspring declines. This suggests that in female humans, selection for deferred ageing should wane when one's own offspring become post-reproductive and, correspondingly, we show that rates of female mortality accelerate as their offspring terminate reproduction.

Life-history theory generally predicts that there should be no selection for living beyond one's reproductive capacity^{1,2}. That female humans show a prolonged post-reproductive lifespan is therefore puzzling. The most compelling hypothesis so far predicts that selection will favour a prolonged post-reproductive lifespan if this enables individuals to increase their fitness through assisting their own offspring to reproduce successfully (the grandmother hypothesis)^{3,4}. Such helping by post-reproductive individuals is unusual in the animal kingdom: helpers are typically pre-reproductive offspring that delay their own dispersal and help their parents to breed⁷. Nevertheless, evidence of helper effects from cooperatively breeding animals can be used to predict the benefits that mothers may confer by helping offspring to reproduce. Helpers may 'lighten the reproductive load' of breeders (allowing them to breed earlier and/or more often^{8,9}) or provide care that is additive to breeders (causing survival improvements to their offspring^{9–11}).

Previous research on humans provides some evidence that post-reproductive mothers can benefit the reproductive output of their offspring^{12–15}. However, it has not yet been possible to test whether prolonged post-reproductive longevity in humans is associated with greater grandchild production, and hence greater fitness, because data covering the complete reproductive histories of several generations of individuals are scarce. Here we investigate the fitness benefits of prolonged post-reproductive lifespan using complete, multi-generational individual-based data sets from humans living in two different countries (Finland, $n = \sim 500$ women and Canada, $n = \sim 2,300$ women) during the eighteenth and nineteenth centuries (see Methods). We concentrate on the fitness benefits of post-reproductive lifespan for females, because the post-reproductive lifespan of males cannot easily be defined. Our data come from historical farming communities, where grandparent(s) are known to have been an integral part of the family, residing in the same house as at least one of their offspring and near to virtually all others^{16,17}. Evidence from contemporary human populations shows that grandparents may assist philopatric offspring by transferring knowledge and participating in household tasks and child care, and that such help may increase offspring breeding probability¹², and grandchildren nutrition and survival^{4,14}.

We first investigate the effect of a woman's post-reproductive lifespan (age at death after ~ 50 , see Methods) on the number of grandchildren that she leaves in the population (that is, fitness). Second, using the Finnish data set, we investigate the effect that the presence of a post-reproductive mother has on the lifetime reproductive performance of all her offspring: lifetime fecundity, lifetime reproductive success (number of children produced and raised to 15) and individual λ ¹⁸ (a fitness measure incorporating the timing of reproduction and the lifetime reproductive output) (see Methods). Third, we investigate the mechanism(s) through which post-reproductive females influence the fitness of their offspring (and hence themselves) by comparing the effect of post-reproduc-

tive mother presence/absence on offspring key life-history traits (age at first reproduction, inter-birth intervals and reproductive tenure length) and grandchildren survival probability. Analyses are conducted using general linear (mixed) models (GLMs or GLMMs) in which we control for a large range of potentially confounding effects (see Methods). These include socio-economic status (rich, wealthy, average, poor), temporal (birth cohort) and geographic (population) differences in living conditions¹⁹, post-reproductive mother's age (or potential age had she been alive), offspring sex, number of competing siblings, and birth order, as well as the inclusion of different offspring from the same mother.

First, the length of a woman's post-reproductive lifespan has a significant positive effect on the number of grandchildren that she gives rise to in both Finns and Canadians (Fig. 1). In each country, the strength of this relationship is similar, and equates to the equivalent of post-reproductive women gaining two extra grandchildren for every ten years that they survive beyond age 50. This similarity arises despite the social, cultural and life-history differences between the two countries (Table 1 and Methods).

These findings are not confounded by a long post-reproductive lifespan being related to the production of more offspring before age 50. No relationship occurs between the number of offspring delivered by a female in her lifetime and her post-reproductive longevity (Finland: $F_{1,351} = 0.04$, $P = 0.84$; Canada: $F_{1,2360} = 0.11$, $P = 0.74$), and the effect of post-reproductive longevity on number of grandchildren remains significant after controlling for the number of offspring produced (Finland: $\beta = 0.18 \pm 0.05$, $F_{1,325} = 15.29$, $P = 0.0001$; Canada: $\beta = 0.20 \pm 0.04$, $F_{1,2357} = 23.68$, $P < 0.0001$). Moreover, our results are unlikely to be

confounded by between-family differences in living conditions or health. Significant effects of socio-economic status as well as temporal and geographical differences in living conditions on grandchild numbers are controlled (Finnish analysis: $F_{3,326} = 4.97$, $P = 0.0022$; $F_{4,326} = 4.82$, $P = 0.0009$; $F_{4,326} = 14.57$, $P < 0.0001$, respectively). Neither socio-economic status, nor temporal or geographic differences in living conditions influences female post-reproductive lifespan (all values $P > 0.1$). The probability of a Finnish mother dying of the most common cause of death, an infectious disease, is unrelated to either the probability of her offspring ($n = 574$) dying of an infectious disease (GLMM: $\chi^2_1 = 0.33$, $P = 0.56$) or the fecundity of her offspring (GLMM: $\chi^2_1 = 0.09$, $P = 0.77$).

Second, further analysis of the Finnish data set reveals that post-reproductive mothers living when their offspring begin to reproduce have a significant positive effect on the three components of fitness measured. In the presence of a living post-reproductive mother, both sons and daughters have a greater number of offspring in their lifetime (Fig. 2a), and raise more to adulthood (age 15) (that is, greater lifetime reproductive success) (Fig. 2b). Additionally, a living post-reproductive mother has a significant positive effect on her offspring's individual λ^{18} (GLMM: $F_{1,524} = 4.22$, $P = 0.04$). Finally, the longer a post-reproductive mother survives after the offspring started reproducing (controlling for mother's age) the greater her offspring's lifetime reproductive success (GLMM: $F_{1,281} = 8.42$, $P = 0.004$). Previous studies have suggested that post-reproductive mothers might direct their care primarily towards daughters^{3–4}, but our results imply that sons and daughters benefit similarly from the presence of a post-reproductive mother. This is likely to arise in our study populations because both sexes usually remained in their natal population and resided near to their mothers.

Again, our results are unlikely to be confounded by either genetic or non-genetic sources of variation (see above). (1) We control for individual differences in socio-economic status, temporal and geographic differences in living conditions, offspring sex, number of competing siblings and birth order (see Methods); (2) we used a mixed modelling framework, which enables comparisons of the success of different offspring from the same family, when their mother was alive versus when she was dead, while controlling for significant within-family effects ($P = 0.05–0.001$)²⁰; (3) further comparison of the reproductive success of offspring who lived in the same village as their post-reproductive mother versus those that lived elsewhere (>20 km away) reveals that those living near to their mother produce more offspring (Fig. 2c). These analyses therefore provide convincing evidence that it is the presence of a post-reproductive mother in itself that causes offspring to reproduce more successfully.

Third, we found that these fitness effects result from the beneficial effects that a living post-reproductive mother has on (1) the key-life-history traits of her offspring; and (2) the survival of her

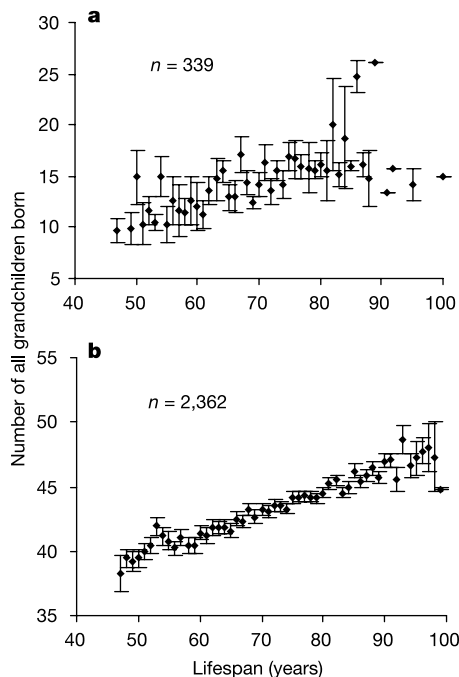


Figure 1 Female lifespan and total number of grandchildren contributed to the following generation in Finland and Canada. **a**, Finland (GLM: $\beta = 0.18 \pm 0.05$, $F_{1,326} = 14.32$, $P = 0.0002$); **b**, Canada (GLM: $\beta = 0.18 \pm 0.05$, $F_{1,2358} = 13.67$, $P = 0.0002$). Graphs show predicted means (± 1 s.e.) after controlling for effects of socio-economic status, and geographic (Finland) and temporal differences in living conditions (Finland and Canada).

Table 1 Life-history parameters of the pre-modern post-reproductive women studied

	Finland			Canada		
	Mean	s.d.	Range	Mean	s.d.	Range
Age at first reproduction (years)	25.4	4.6	16–43	22.8	4.3	15–44
Age at last reproduction (years)	39.3	5.0	19–52	38.7	6.1	16–50
Number of offspring born	6.8	3.0	1–18	9.1	3.9	1–22
Number of adult offspring	3.8	2.2	0–12	5.1	2.9	1–16
Number of grandchildren	11.3	10.8	0–53	38.2	28.0	0–157
Longevity (for those surviving to 50)	67.5	10.2	50–100	74.0	10.9	50–100

Number of adult offspring is the number of offspring raised to age 15 in Finland and the number of offspring surviving to marry in Canada. s.d., standard deviation.

grandchildren. In the presence of a living mother, offspring reproduce earlier, with the difference in ages at first reproduction in the presence and absence of a post-reproductive mother being 2.4 yr (Fig. 2d). This effect is evident after controlling for mother's age (or potential age if dead), and hence is not confounded by the possibility that offspring who begin to breed earlier are more likely

to have a younger, living mother. Furthermore, in the presence of a living mother, offspring have significantly reduced inter-birth intervals (first three intervals only; Fig. 2e), and increased reproductive tenures (time between first and last childbirth) (GLMM: $F_{1,534} = 4.5$, $P = 0.035$). These effects do not differ according to offspring gender, again indicating that such benefits are common to both sons and daughters.

In addition, grandchildren have significantly higher survival probabilities if their grandmother is alive at their birth (Fig. 2f). However, the significance of this effect depends on both the age of the grandmother and the grandchild. Grandchild survival to adulthood is enhanced by 12% when grandmothers are under 60 at their birth, but by only 3% when grandmothers are over this age. Grandmothers have no effect on the survival probability of their grandchildren between birth and two years of age (GLMM: $\chi^2_1 = 0.08$, $P = 0.78$), but have significant positive effects when offspring are between two and five (GLMM: $\chi^2_1 = 7.09$, $P = 0.008$), and between five and 15 (GLMM: $\chi^2_1 = 4.18$, $P = 0.041$). That offspring survival is unaffected by grandmother presence pre-weaning (birth to two years), when offspring are suckled by their own mother, but is affected thereafter, provides strong evidence that the grandchild survival effect observed is again caused by the grandmother in itself, and not by differences in individual quality, or maternal or environmental conditions.

In conclusion, our results lend strong support for the hypothesis that prolonged female post-reproductive lifespan is adaptive^{3,4}, to our knowledge revealing for the first time the substantial fitness benefits that females accrue by living beyond reproductive age. However, the results of this study have broader implications regarding the evolution of cooperative breeding²⁰, menopause⁵ and senescence⁶. First, as this study measures the effects of helping on an individual's complete contribution to a following generation, it represents one of the most detailed studies of fitness conducted in a cooperative breeding vertebrate. Second, although grandmother effects alone are suggested to be insufficient to account for the evolution of menopause⁵, our results suggest that they can be sufficient to account for the evolution of the substantially prolonged post-reproductive lifespan observed in humans²¹.

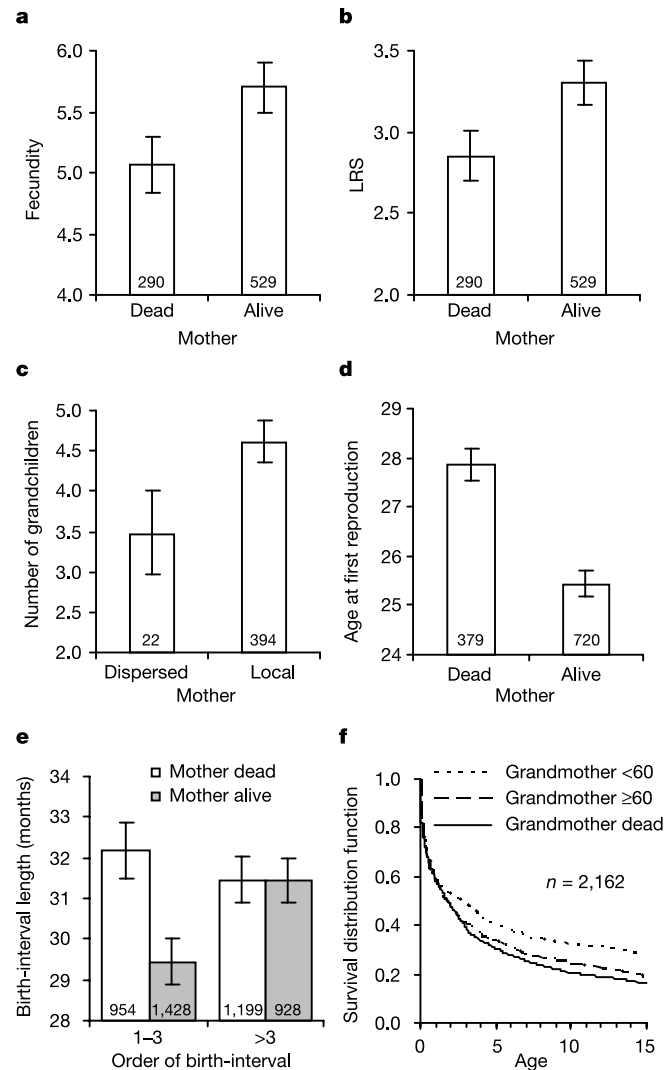


Figure 2 Post-reproductive mother presence (at the time that each offspring began to reproduce) and correlates of each offspring's (son and daughter) fitness and key life-history traits in Finland. **a–c**, Offspring fitness; **d–f**, Key life-history traits. A living post-reproductive mother is associated with: **a**, increased lifetime fecundity of offspring ($F_{1,607} = 9.44$, $P = 0.0022$); **b**, increased lifetime reproductive success of offspring (LRS) ($F_{1,593} = 8.23$, $P = 0.0043$); **c**, increased numbers of grandchildren produced when post-reproductive mother was living in the same village versus elsewhere ($F_{1,293} = 3.98$, $P = 0.047$); **d**, reduced ages at first reproduction of offspring ($F_{1,814} = 52.65$, $P < 0.0001$); **e**, reduced first three inter-birth intervals ($F_{1,977} = 6.45$, $P = 0.011$), but not those thereafter ($F_{1,711} = 0.84$, $P = 0.36$); and **f**, increased survival of grandchildren until age 15 (log-rank test: $\chi^2_2 = 19.78$, $P < 0.0001$). Panels **a–e** show predicted means (± 1 s.e.) from GLMMs after controlling for effects of offspring socio-economic status, geographic and temporal differences in living conditions, birth order, sex, number of siblings, mother's age and repeated measures of mother, where appropriate. Panel **f** shows Kaplan–Meier survival curves for grandchildren depending on the age and presence of a post-reproductive grandmother.

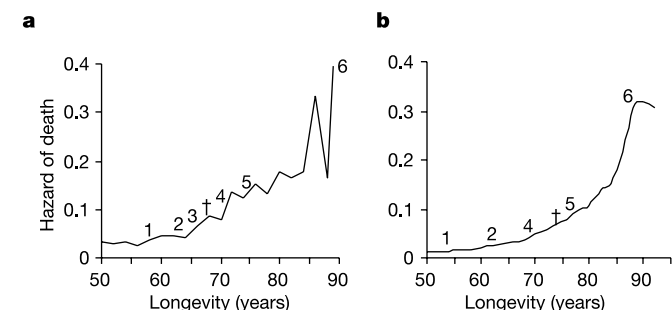


Figure 3 Age-specific hazards of death for post-reproductive mothers and the timing of different life-history events of their offspring and grandchildren. **a**, In Finland; **b**, in Canada. Dagger indicates average age at death post-reproduction. 1, average age when offspring have first offspring; 2, average age when offspring have their fourth offspring; 3, maximum age when grandchild survival can be significantly improved; 4, average age when offspring produce their last offspring; 5, average age when last offspring have their last offspring; 6, average age when 50% of all grandchildren start producing great-grandchildren. Note that female mortality rates are relatively low when most increments to fitness can be gained through helping offspring but high when fitness could be gained through helping grandchildren, to whom they are half as related.

Finally, our results not only show that selection may act to increase post-reproductive longevity, but also indicate at what age such selection might be expected to wane (Fig. 3). The mean lifespan of post-reproductive women from Finland and Canada was 68 and 74 yr respectively, which is close to the maximum age at which women can gain fitness in the two countries and the age at which offspring cease reproducing. That mortality rates accelerate from the time that offspring begin to terminate reproduction (Fig. 3) suggests that selection for post-reproductive longevity is deferred only until a woman's own offspring finish reproducing and can become the future generation of helpers. □

Methods

The Finnish women

The Finnish data were collected using historical population registers. The Lutheran Church has been obliged by law to submit accurate census registers of all births, marriages and deaths in each parish in the country since the seventeenth century²², making the Finnish church registers one of the most reliable sources of demographic data on historical humans²³. Our data contain two complete generations of pedigree data for 537 reproductive women from five farming/fishing communities, varying substantially in their ecological conditions¹⁹.

We recorded the survival and lifetime reproductive details (Table 1) of 537 women born during the years 1702–1823 ($n = 3,629$ offspring born, 1,186 marrying and reproducing in the population, 6,002 grandchildren born). The study period therefore ended before industrialism and more liberal economics, healthcare improvements and modern birth-control methods influenced standard of living, survival and reproduction²⁴. The occupation of each woman's husband (at the time the children were born, for example, priest, landowner, tenant farmer, servant, and so on) allowed us to rank the socio-economic status (rich, wealthy, average or poor) of each family, which is a correlate of resources available²⁵.

The Canadian women

The Canadian data were obtained using the BALSAC population register based at the University of Quebec, Chicoutimi²⁶. The register, internationally recognized for its quality, contains demographic and genealogical information collected from baptism, marriage and death certificates from all individuals in the Saguenay region of Quebec during the nineteenth and twentieth centuries²⁶. The population is mostly French-speaking and Catholic, and until the beginning of the twentieth century, mainly agricultural. This study includes all women ($n = 3,290$) born in the Saguenay region from 1850 to 1879 (inclusive) who got married in the region and had at least one offspring who also married in the region ($n = 29,895$ children born, 16,618 marrying and reproducing in the population, 100,074 grandchildren born; Table 1; ref. 27). Detailed data on the reproductive history of each woman, other than the number, survival and subsequent reproductive success of their offspring, were unavailable to us, enabling us to conduct the analyses shown in Fig. 2 for Finns only.

Statistical analyses

Statistical analyses were conducted using SAS (SAS Institute Inc., release 8.02, 1999–2001). Relationships between a female's post-reproductive lifespan and the number of grandchildren born to her were analysed using general linear models, in which number of grandchildren born was fitted as the response term to a normal error structure and female longevity was fitted as the main fixed effect. In these analyses, we restricted our data to include only those women who lived past menopause ($n = 339$ for Finland and $n = 2,362$ for Canada)—the age at which 99% of females had finished reproducing (~ 49 yr). The analyses control for socio-economic status and geographic differences in living conditions (population)¹⁹ in Finland and temporal differences in living conditions (birth cohort in both countries).

Effects of post-reproductive mother presence on fitness correlates (Fig. 2a–c) and underlying life-history traits of offspring (Fig. 2d, e) were analysed using GLMMs, which allow both fixed and random terms to be fitted to a model, with random terms controlling for repeated measures within mother²⁸ (and grandmother in Fig. 2e). Satterthwaite's formula²⁹ was used to approximate the denominator degrees of freedom of each fixed effect. In these analyses, the response terms were fitted to a normal error structure and mother presence/absence was fitted as the main fixed effect. We defined presence/absence as whether a post-reproductive mother was alive versus dead (Fig. 2a, b, d), or living in the same village versus elsewhere (Fig. 2c), when each offspring started reproducing. In Fig. 2e, the presence of a post-reproductive (grand)mother was defined as whether or not she was alive when the grandchild was born. We controlled for confounding effects (where appropriate) of: mother's age, offspring's socio-economic status, population¹⁹, birth cohort, sex, birth order, the number of competing siblings, the survival status of the directly previous child born in the family and the repeated measures of the same mother.

The effect of post-reproductive grandmother presence on grandchild survival was analysed using log-rank test and survivorship functions estimated using the Kaplan–Meier method. To investigate the period when the presence of the post-reproductive woman was most important for the survival of grandchildren, GLMMs were performed, where the response term for whether or not the grandchild survived to a certain age was fitted to a

binomial error structure and logit-link function. Analyses controlled for socio-economic status of a parent, population, birth cohort and order, sex of grandchildren and repeated measures within parents.

In all models, significant model terms were determined using the backward elimination technique, and subsequently checked using the forward technique. The residuals of all models were normally distributed (GLMs, GLMMs) and the variances were equal (GLM: Levene's test $P > 0.05$). All statistics are from the GLM unless otherwise indicated.

Received 5 January; accepted 23 January 2004; doi:10.1038/nature02367.

- Williams, G. C. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* **11**, 398–411 (1957).
- Hamilton, W. D. The moulding of senescence by natural selection. *J. Theor. Biol.* **12**, 12–45 (1966).
- Hawkes, K., O'Connell, J. F., Blurton Jones, N. G., Alvarez, H. & Charnov, E. L. Grandmothering, menopause, and the evolution of human life histories. *Proc. Natl Acad. Sci. USA* **95**, 1336–1339 (1998).
- Hawkes, K., O'Connell, J. F. & Blurton Jones, N. G. Hadza women's time allocation, offspring provisioning, and the evolution of long postmenopausal lifespans. *Curr. Anthropol.* **38**, 551–577 (1997).
- Shanley, D. P. & Kirkwood, T. B. L. Evolution of the human menopause. *Bioessays* **23**, 282–287 (2001).
- Lee, R. D. Rethinking the evolutionary theory of aging: Transfers, not births, shape senescence in social species. *Proc. Natl Acad. Sci. USA* **100**, 9637–9642 (2003).
- Emlen, S. T. An evolutionary theory of the family. *Proc. Natl Acad. Sci. USA* **92**, 8092–8099 (1995).
- Russell, A. E., Brotherton, P. N. M., McIlrath, G. M., Sharpe, L. L. & Clutton-Brock, T. H. Breeding success in cooperative meerkats: effects of helper number and maternal state. *Behav. Ecol.* **14**, 486–492 (2003).
- Brown, J. L., Brown, E. R., Brown, S. D. & Dow, D. D. Helpers: effects of experimental removal on reproductive success. *Science* **215**, 421–422 (1982).
- Emlen, S. T. & Wrege, P. H. Breeding biology of white-fronted bee-eaters at Nakuru: the influence of helpers on breeder fitness. *J. Anim. Ecol.* **60**, 309–326 (1991).
- Clutton-Brock, T. H. *et al.* Effects of helpers on juvenile development and survival in meerkats. *Science* **293**, 2446–2449 (2001).
- Sear, R., Mace, R. & McGregor, I. A. The effects of kin on female fertility in rural Gambia. *Evol. Hum. Behav.* **24**, 25–42 (2003).
- Voland, E. & Beise, J. Opposite effects of maternal and paternal grandmothers on infant survival in historical Krummhörn. *Behav. Ecol. Sociobiol.* **52**, 435–443 (2002).
- Sear, R., Mace, R. & McGregor, I. A. Maternal grandmothers improve the nutritional status and survival of children in rural Gambia. *Proc. R. Soc. Lond. B* **267**, 1641–1647 (2000).
- Jamison, C. S., Cornell, L. L., Jamison, P. L. & Nakazato, H. Are all grandmothers equal? A review and a preliminary test of the grandmother hypothesis in Tokugawa Japan. *Am. J. Phys. Anthropol.* **119**, 67–76 (2002).
- Moring, B. Household and family in Finnish coastal societies 1635–1895. *J. Fam. Hist.* **18**, 395–414 (1993).
- Bouchard, G. *Quelques Arpents d'Amérique. Histoire, Population et Famille au Saguenay, 1838–1971* (Boréal, Montréal, 1996).
- McGraw, J. B. & Caswell, H. Estimation of individual fitness from life-history data. *Am. Nat.* **147**, 47–64 (1996).
- Lummaa, V., Haukioja, E., Lemmetyinen, R. & Pikkola, M. Natural selection on human twinning. *Nature* **394**, 533–534 (1998).
- Cockburn, A. Evolution of helping behavior in cooperatively breeding birds. *Annu. Rev. Ecol. Syst.* **29**, 141–177 (1998).
- Hawkes, K. Grandmothers and the evolution of human longevity. *Am. J. Hum. Biol.* **15**, 380–400 (2003).
- Luther, G. *Suomen Tilastotoimen Historia vuoteen 1970* (WSOY, Helsinki, 1993).
- Gille, H. The demographic history of the northern European countries in the eighteenth century. *Popul. Stud.* **3**, 3–65 (1949).
- Soininen, A. M. *Old Traditional Agriculture in Finland in the 18th and 19th Centuries* (Forssan Kirjapaino Oy, Forssa, 1974).
- Karskela, S. *Sukututkijan Tietokirja* (Gummerus Kirjapaino Oy, Saarijärvi, 2001).
- Bouchard, G., Roy, R., Casgrain, B. & Hubert, M. in *From Information to Knowledge: Conceptual and Content Analyses by Computer* (eds Nissan, E. & Schmidt, K.) 201–227 (Intellect, Oxford, 1995).
- Lummaa, V. & Tremblay, M. Month of birth predicted reproductive success and fitness in pre-modern Canadian women. *Proc. R. Soc. Lond. B* **270**, 2355–2361 (2003).
- Schall, R. Estimation of generalized linear models with random effects. *Biometrika* **78**, 719–727 (1991).
- Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. *SAS System for Mixed Models* (SAS Institute Inc, Cary, 1996).

Acknowledgements We are grateful to T. Clutton-Brock, N. Davies, B. Hatchwell, E. Haukioja, J. Jokela and R. Kilner for comments, T. Coulson for help with calculating individual λ values, and K. Pokkinen, A. Siitonen and T. Verho for collecting the Finnish demographic data. We also thank the Royal Society, UK (V.L. and A.F.R.), Academy of Finland (V.L.), Wihuri Foundation (M.L.), Emil Aaltonen Foundation (S.H.) and Social Sciences and Humanities Research Council of Canada (M.T.) for funding.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.L. (mipayl@utu.fi).

Immunogenicity of a highly attenuated MVA smallpox vaccine and protection against monkeypox

Patricia L. Earl¹, Jeffrey L. Americo¹, Linda S. Wyatt¹, Leigh Anne Eller², J. Charles Whitbeck³, Gary H. Cohen³, Roselyn J. Eisenberg³, Christopher J. Hartmann⁴, David L. Jackson⁴, David A. Kulesh⁴, Mark J. Martinez⁴, David M. Miller⁴, Eric M. Mucker⁴, Joshua D. Shamblin⁴, Susan H. Zwiars⁴, John W. Huggins⁴, Peter B. Jahrling⁴ & Bernard Moss¹

¹National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0445, USA

²Henry M. Jackson Foundation, Rockville, Maryland 20850, USA

³Schools of Dental and Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

⁴United States Army Medical Research Institute of Infectious Diseases, Frederick, Maryland 21702-5011, USA

The potential use of smallpox as a biological weapon has led to the production and stockpiling of smallpox vaccine and the immunization of some healthcare workers. Another public health goal is the licensing of a safer vaccine that could benefit the millions of people advised not to take the current one because they or their contacts have increased susceptibility to severe vaccine side effects¹. As vaccines can no longer be tested for their ability to prevent smallpox, licensing will necessarily include comparative immunogenicity and protection studies in non-human primates. Here we compare the highly attenuated modified vaccinia virus Ankara (MVA)² with the licensed Dryvax vaccine in a monkey model. After two doses of MVA or one dose of MVA followed by Dryvax, antibody binding and neutralizing titres and T-cell responses were equivalent or higher than those induced by Dryvax alone. After challenge with monkeypox virus, unimmunized animals developed more than 500 pustular skin lesions and became gravely ill or died, whereas vaccinated animals were healthy and asymptomatic, except for a small number of transient skin lesions in animals immunized only with MVA.

MVA was generated by more than 500 passages of vaccinia virus in chick embryo fibroblasts, during which it acquired multiple deletions and mutations and lost the capacity to replicate efficiently in human and most other mammalian cells²⁻⁷. The host restriction of MVA occurs during a late stage of viral morphogenesis, resulting in the accumulation of immature virions and the formation of abnormal ones^{5,8}. As a result of extreme attenuation, MVA causes no adverse effects even when high doses are injected into immune-deficient non-human primates⁹. Animal studies carried out more than 30 years ago showed that MVA protected animals against orthopoxvirus infections, although few immune responses were measured and those that were measured appeared low^{10,11}. In a recent study, MVA and the smallpox vaccine induced a similar variety of immune responses in mice¹². MVA seemed to be safe in humans and there was a reduction in the size of skin lesions produced by a subsequent standard smallpox vaccination¹³⁻¹⁵. Monkeypox currently provides the best non-human primate model of an orthopoxvirus infection^{16,17}. Monkeypox is endemic in the rainforests of central and western Africa¹⁸, although a recent outbreak occurred in the United States. Human monkeypox is an exanthematous disease that can resemble smallpox in the severity of its symptoms and could also be a potential biological weapon¹⁹.

MVA is being considered as a replacement for the present smallpox vaccine for those with a high risk of adverse complications, or for more general use as a pre-vaccine. In the first case, the

immune responses elicited by one or two doses of MVA should approach although not necessarily equal those of the licensed smallpox vaccine. In the second, MVA should reduce the reaction to a subsequent smallpox vaccine without diminishing the total immune response. For immunization of monkeys, we decided on an inoculum of 10⁸ plaque-forming units (PFU). This was based on mouse experiments showing that the antibody response to MVA increased with dose and that at 10⁸ PFU equalled or exceeded the response to Dryvax (L. Wyatt, unpublished observations). Twenty-four cynomolgous monkeys were divided into four groups: group 1 received an intramuscular inoculation with 10⁸ PFU of MVA at zero time and a second two months later; group 2 received one intramuscular injection with 10⁸ PFU of MVA followed two months later by a standard percutaneous inoculation with Dryvax; group 3 received nothing at zero time and one Dryvax inoculation two months later; group 4 served as the unimmunized control. No adverse local or systemic effects were noted after vaccination with MVA. As expected, pustular skin lesions developed after Dryvax. When Dryvax was given alone, the lesions reached maximal size on days 7 to 10 and persisted for a further one to two weeks before the scabs fell off (Fig. 1). The lesions were smaller, less indurated and healed more rapidly when MVA preceded Dryvax (Fig. 1 and Supplementary Fig. 1), providing initial evidence for the immunogenicity of MVA in this model and its ability to attenuate the effects of a subsequent smallpox vaccination.

After the MVA and Dryvax immunizations, we measured the antibody responses of monkeys to the two infectious forms of vaccinia virus. Intracellular mature virions (IMV) are liberated by cell lysis and are thought to be responsible for host-to-host spread because of their stability in the environment. Extracellular enveloped virions are essentially IMV surrounded by an additional fragile membrane and have an important role in dissemination within the host²⁰. Importantly, the two infectious forms have different outer membrane proteins, and antibodies to those of each membrane contribute to protection²¹⁻²⁴. We first measured antibody titres to purified intact IMV. The geometric mean averages are shown in Fig. 2a and the titres of individual animals in Supplementary Fig. 2a. The response to the first MVA inoculation was detected at one week, peaked at two to four weeks, and was boosted one week after the second MVA dose or two weeks after Dryvax. The single Dryvax immunization elicited a maximal antibody response after four weeks (Fig. 2a and Supplementary Fig. 2a). The additional week

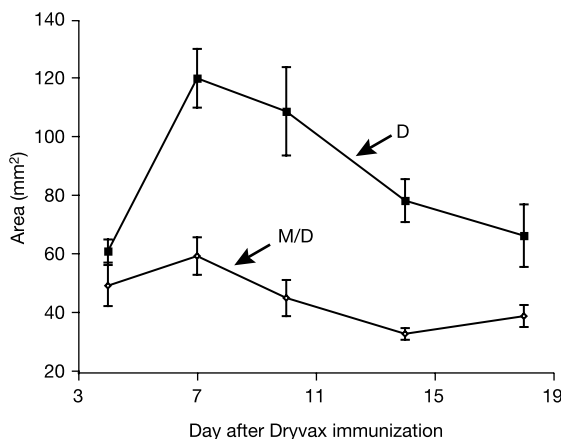


Figure 1 Sizes of Dryvax-induced lesions. Twelve monkeys were immunized percutaneously with Dryvax. Six of the monkeys (M/D) had been immunized intramuscularly with MVA two months earlier and six (D) had no prior exposure to MVA. The area of induration at the site of Dryvax inoculation was measured every three to four days. Average values are given with standard errors.

taken for Dryvax to induce a response, compared with MVA, presumably reflects the need for the low Dryvax inoculum (approximately 2.5×10^5 PFU) to replicate in the skin. By contrast, the replication-defective MVA was administered as a bolus of 10^8 PFU. The maximal antibody responses after single MVA and Dryvax inoculations were not significantly different. Similarly, the responses after the MVA and Dryvax boosts were similar, each inducing a tenfold higher maximal antibody titre than the single Dryvax immunization. The L1R protein, a component of the IMV membrane, is a target of neutralizing antibody and has been shown to induce partial protection in mouse models^{22,25}. The availability of a soluble recombinant L1R protein reactive with neutralizing monoclonal antibodies (to be described elsewhere) allowed us to determine the antibody responses to this important antigen. The kinetics of L1R antibody formation generally followed that found for whole IMV (Fig. 2b).

Next, we determined antibody titres to two membrane proteins of extracellular enveloped virions, A33R and B5R, each of which

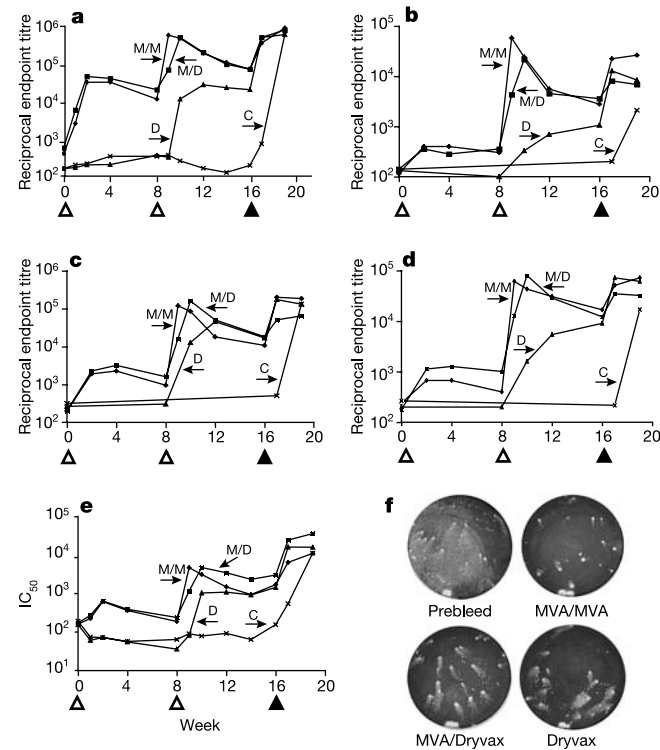


Figure 2 Binding and vaccinia virus neutralizing antibody responses. **a–d**, Enzyme-linked immunosorbent assay (ELISA) titres were determined using microtitre plates coated with 10^6 PFU per well of sucrose gradient purified vaccinia virus strain WR and fixed with paraformaldehyde (**a**), 0.04 μ g of purified L1R protein (**b**), 0.09 μ g of purified A33R protein (**c**) or 0.1 μ g of purified B5R protein (**d**). Sera were heat inactivated and endpoint titres were determined as the maximum serum dilution at which the absorbance was >0.10 after subtraction of background. Geometric mean ELISA titres of the six monkeys in each group are shown. **e**, Neutralizing antibody titres were determined using a flow cytometric assay that measures reduction in infectivity of a virus expressing GFP²⁶. All assays were performed in duplicate. Geometric mean titres are shown. **f**, Reduction of cell-to-cell spread of vaccinia virus. Serum samples from antibody peak time points (MVA/MVA, week 9; MVA/Dryvax, week 10; Dryvax, week 12) or before immunization (prebleed) diluted 1:66 were added to the medium following inoculation of BS-C-1 cell monolayers with a vaccinia virus strain, IHD-J, which produces large amounts of released extracellular virus. After 48 h, the monolayers were stained with crystal violet. Open triangles, times of immunization with MVA or Dryvax; filled triangles, time of challenge with monkeypox. C, unimmunized control; M/M, MVA prime at week 0 and boost at week 8; M/D, MVA prime at week 0 and Dryvax boost at week 8; D, Dryvax at week 8.

has induced active and passive protective immunity in mouse models^{22–24}. The availability of soluble recombinant proteins that react with conformational monoclonal antibodies (to be described elsewhere) again allowed us to determine specific antibody responses. Although responses to both proteins were found after single MVA and Dryvax immunizations, higher titres were achieved with the latter (Fig. 2c, d). Nevertheless, a second MVA or MVA followed by Dryvax boosted the A33R and B5R antibody titres to even higher levels than those achieved with one Dryvax immunization.

Neutralization titres were determined by incubating purified recombinant IMV with serum dilutions, infecting cells and measuring the reduction in the percentage of cells that expressed virus-

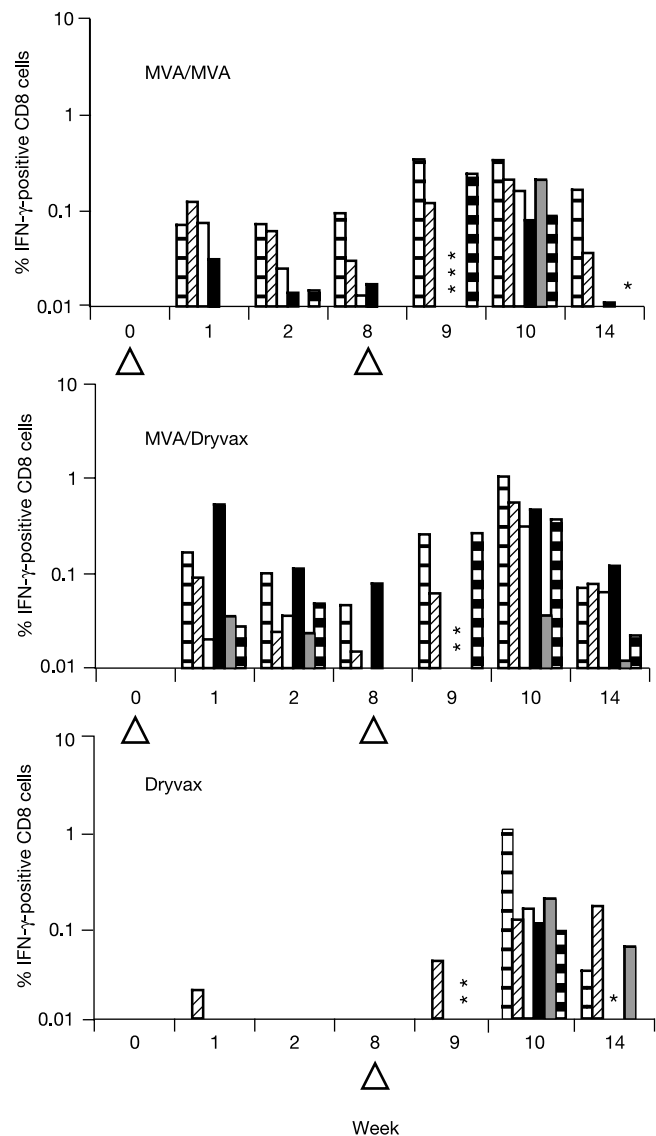


Figure 3 Induction of vaccinia-virus-specific IFN- γ -producing T cells. Fresh peripheral blood mononuclear cells from monkeys were infected with 6 PFU of vaccinia virus (strain WR) per cell overnight in the presence of Brefeldin A essentially as described³⁰. Antigens were stained with the following anti-human antibodies: CD3 (phycoerythrin), CD8 (fluorescein isothiocyanate) and IFN- γ (allophycocyanin) (BD Pharmingen), enumerated with a FACScaliber and analysed with FloJo software. Percentage of CD8 T cells that expressed IFN- γ are shown for individual animals. *Data not available. Open triangles, times of immunization.

Table 1 Clinical signs of monkeypox disease

Clinical signs	Controls	MVA/MVA	MVA/Dryvax	Dryvax
Temperature	>103°F on day 3	Normal	Normal	Normal
Weight	4–10% weight loss	Normal	Normal	Normal
Activity	Sluggish by day 9–12, moribund by day 15–18	Normal	Normal	Normal
Pox lesions per animal	> 500 each	1, 1, 11, 16, 32, 36	None	None
Onset of pox lesions	Day 3–6	Day 9–15	None	None
Quality of pox lesions	Large, pustular, typical progression	Small, atypical, non-progressive	NA	NA
Resolution of pox lesions	Not resolved	Within 3–9 days	NA	NA
Other clinical signs	Required subcutaneous fluids and force feeding	None	None	None
Death	Two by day 18	None	None	None

Clinical signs were monitored daily. NA, not applicable.

encoded enhanced green fluorescent protein (GFP) by flow cytometry²⁶. Neutralizing antibodies were detected after the initial immunizations with MVA or Dryvax, although the titres were higher with the latter vaccine (Fig. 2e and Supplementary Fig. 2b). Nevertheless, MVA priming followed by either MVA or Dryvax raised higher neutralization titres than did Dryvax alone. At the time of challenge with monkeypox, however, there was no statistical difference between the groups.

Antibodies that prevent the spread of extracellular vaccinia virus can be detected by a comet reduction assay in which antiserum is added to the liquid overlay of a cell monolayer after virus infection²⁷. The comets represent satellite plaques that are caused by the spread of released virus by convection currents²⁸. Sera from animals before immunization had no effect on the appearance of the diffuse comets, which nearly covered the cell monolayer, whereas sera from immunized animals greatly reduced their size. Representative results are shown in Fig. 2f and those from sera of all animals are shown in Supplementary Fig. 3.

Specific T-cell responses may also have a role in protection against orthopoxvirus infections and were measured by immunostaining for interferon- γ (IFN- γ) (Fig. 3). CD8⁺ T-cell responses to vaccinia virus were detected at one week after MVA immunization (week 0) and two weeks after Dryvax (week 8). All animals made specific responses, but there was considerable individual variation, and the differences between first and second vaccinations and between immunized groups were not statistically significant. CD4 responses were lower than CD8 responses, but were positive for the majority of animals in each immunization group (data not shown).

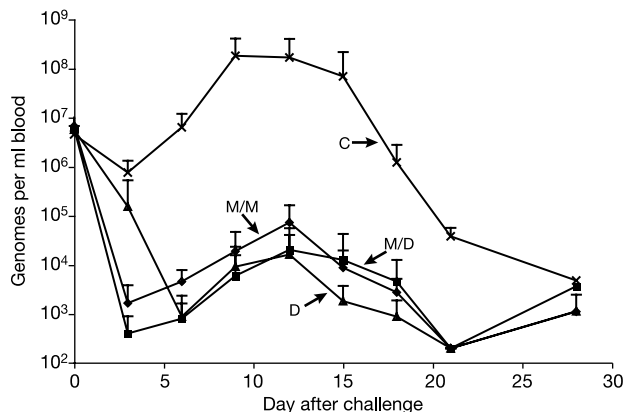


Figure 4 Viral load following monkeypox virus challenge. Number of monkeypox genomes per ml of blood was determined by extraction of DNA with the QIAGEN QIAamp DNA Mini Kit and quantitative TaqMan-MGB PCR using a pan-orthopoxvirus probe and primer set directed against the viral haemagglutinin gene. The limit of quantification from whole blood is 5×10^3 genomes or 100 PFU. Average values for each group with standard deviations are shown. Two animals from the control group died between days 15 and 18.

The normal route of smallpox transmission is through the mouth, nose and pharynx²⁹. After a latent phase of up to two weeks, viraemia occurs, resulting in the seeding of organs including skin. Because a consistent upper respiratory tract monkeypox model is not available, we used intravenous injection, which effectively bypasses the incubation and prodromal phases of naturally acquired orthopoxviral infection in humans, by creating instantaneous viraemia and systemic spread of virus to target tissues. Intravenous inoculation is probably the most rigorous challenge route. Two months after the last immunization, each monkey was challenged by intravenous injection of 5×10^7 PFU of monkeypox virus. Binding and neutralizing antibody titres rose rapidly in all of the immunized groups and more slowly in the previously unimmunized animals (Fig. 2 and Supplementary Fig. 2). Because of containment issues, we were unable to measure the T-cell responses following challenge. Vaccinated animals had low levels of monkeypox viral DNA in the blood as measured by quantitative TaqMan 3'-Minor Groove Binder (MGB) polymerase chain reaction (PCR) (Fig. 4). These levels were four logs lower than those in unvaccinated controls, with no significant difference between vaccine groups.

Although the PCR data suggested a low viraemia, all immunized animals remained clinically well, whereas unimmunized animals became extremely ill with fever, weight loss and diminished activity (Table 1). Despite supportive care including subcutaneous fluids, one unimmunized animal died and another had to be euthanized. The unimmunized monkeys each developed 500 to 1,000 or more cutaneous pox lesions that exhibited a pustular progression and had not resolved when the experiment was terminated at 30 days after the challenge. Monkeys immunized with Dryvax or with MVA and Dryvax had no detectable lesions. Two of the monkeys immunized with MVA alone developed only a single lesion and the others had 11 to 36 lesions each. The lesions in MVA-immunized monkeys appeared later than those in unimmunized animals and resolved faster without going through a pustular stage (Supplementary Fig. 4). There was no correlation between the number of lesions and any of the measured systemic immune responses in the MVA-immunized group. A reduction in lesions was also found when MVA vaccinations were followed by subcutaneous injection of monkeys with variola virus¹¹. We are unsure whether the difference between Dryvax- and MVA-immunized monkeys was due to the vaccines *per se* or to the routes of inoculation—that is, skin versus intramuscular. The induction of skin-homing T cells in the animals inoculated percutaneously with Dryvax is an interesting possibility.

Because the correlates of smallpox protection are unknown, our findings of similar humoral and cellular immune responses to MVA and Dryvax in non-human primates and substantial protection against a severe monkeypox virus challenge are important steps in the evaluation of MVA as a replacement vaccine for those with increased risk of severe side effects from the standard live vaccine, or as a pre-vaccine. One approach would be to vaccinate with MVA before any immediate smallpox threat, with the expectation that a standard vaccine or a second MVA (depending on individual risk

factors) would be given as a boost in the event of documented smallpox. Further experiments to determine the longevity of protection, the consequences of delayed boosting and dosage effects are planned. □

Methods

Virus preparations

A vial of MVA passage 572 (2/22/1974), obtained from A. Mayr, was plaque purified, propagated in chick embryo fibroblasts, and purified by sedimentation through a sucrose cushion (to be described elsewhere). A vial of Dryvax (Wyeth Laboratories, Inc.), obtained from the US Centers for Disease Control and Prevention, was diluted as recommended and stored at 4 °C. Monkeypox virus strain Zaire 79 (V79-I-005), originally isolated from the scab of a fatally infected human in LLC-MK2 cells and passed twice in BSC40 cells, was obtained from J. Esposito and propagated in MA-104 cells. A clarified lysate of infected cells was used to make the challenge stock of virus.

Animals

Cynomolgous monkeys (*Macaca fascicularis*) of Chinese origin were housed at Bioqual, Inc. for the immunological studies and transferred to the US Army Medical Research Institute of Infectious Diseases before the monkeypox virus challenge.

Vaccinia virus neutralization assay²⁶

Twofold serial dilutions of heat-inactivated sera were incubated with virus encoding the GFP for 1 h at 37 °C. HeLa S3 cells were added and incubated overnight. GFP-expressing cells were enumerated with a FACScaliber and analysed with FloJo software. IC₅₀ values were calculated using PRISM software (GraphPad Software, Inc.). All assays were performed in duplicate.

Received 26 November 2003; accepted 7 January 2004; doi:10.1038/nature02331.

- Fulginiti, V. A., Papier, A., Lane, J. M., Neff, J. M. & Henderson, D. A. Smallpox vaccination: a review, part II. Adverse events. *Clin. Infect. Dis.* **37**, 251–271 (2003).
- Mayr, A., Hochstein-Mintzel, V. & Stöckl, H. Abstammung, eigenschaften und verwendung des attenuierten vaccinia-stammes MVA [Passage history, properties, and applicability of the attenuated vaccinia virus strain MVA]. *Infection* **3**, 6–14 (1975).
- Meyer, H., Sutter, G. & Mayr, A. Mapping of deletions in the genome of the highly attenuated vaccinia virus MVA and their influence on virulence. *J. Gen. Virol.* **72**, 1031–1038 (1991).
- Antoine, G., Scheiflinger, E., Dorner, F. & Falkner, F. G. The complete genomic sequence of the modified vaccinia Ankara strain: comparison with other orthopoxviruses. *Virology* **244**, 365–396 (1998).
- Carroll, M. & Moss, B. Host range and cytopathogenicity of the highly attenuated MVA strain of vaccinia virus: propagation and generation of recombinant viruses in a nonhuman mammalian cell line. *Virology* **238**, 198–211 (1997).
- Drexler, I., Heller, K., Wahren, B., Erfle, V. & Sutter, G. Highly attenuated modified vaccinia virus Ankara replicates in baby hamster kidney cells, a potential host for virus propagation, but not in various human transformed and primary cells. *J. Gen. Virol.* **79**, 347–352 (1998).
- Blanchard, T. J., Alcamí, A., Andrea, P. & Smith, G. L. Modified vaccinia virus Ankara undergoes limited replication in human cells and lacks several immunomodulatory proteins: implications for use as a human vaccine. *J. Gen. Virol.* **79**, 1159–1167 (1998).
- Sutter, G. & Moss, B. Nonreplicating vaccinia vector efficiently expresses recombinant genes. *Proc. Natl Acad. Sci. USA* **89**, 10847–10851 (1992).
- Stittelaar, K. J. et al. Safety of modified vaccinia virus Ankara (MVA) in immune-suppressed macaques. *Vaccine* **19**, 3700–3709 (2001).
- Hochstein-Mintzel, V., Huber, H.-C. & Stöckl, H. Die orale und nasale immunisierung mit poxvirus vaccinae III. Mitteilung: tierexperimentelle untersuchungen [Oral and nasal immunization with poxvirus vaccine. Part III: experimental animal investigations]. *Zentralbl. Bakteri. Hyg.* **156**, 30–96 (1972).
- Hochstein-Mintzel, V., Hänichen, T., Huber, H. C. & Stöckl, H. Vaccinia- und variolaprotektive wirkung des modifizierten vaccinia-stammes MVA bei intramuskulärer immunisierung [An attenuated strain of vaccinia virus (MVA). Successful intramuscular immunization against vaccinia and variola]. *Zentralbl. Bakteri. Hyg.* **230**, 283–297 (1975).
- Belyakov, I. M. et al. Shared modes of protection against poxvirus infection by attenuated and conventional smallpox vaccine viruses. *Proc. Natl Acad. Sci. USA* **100**, 9458–9463 (2003).
- Stöckl, H. et al. MVA-stufenimpfung gegen pocken. klinische erprobung des attenuierten pocken-lebendimpfstoffes, stamm MVA [MVA vaccination against smallpox: clinical trials of an attenuated live vaccinia virus strain (MVA)]. *Dtsch. Med. Wochenschr.* **99**, 2386–2392 (1974).
- Mayr, A., Stöckl, H., Müller, H. K., Danner, K. & Singer, H. Der pockenimpfstamm MVA: marker, genetische struktur, erfahrungen mit der parenteralen schutzimpfung und verhalten im abwehrgeschwächten organismus [The smallpox vaccination strain MVA: marker, genetic structure, experience gained with parenteral vaccination and behavior in organisms with debilitated defense mechanism]. *Zentralbl. Bakteri. Hyg.* **167**, 375–390 (1978).
- Mayr, A. History of variola, smallpox-eradication and MVA. *Berl. Munch. Tierarztl. Wochenschr.* **112**, 322–328 (1999).
- Zauch, G. M., Jahrling, P. B., Geisbert, T. W., Swearingen, J. R. & Hensley, L. The pathology of experimental aerosolized monkeypox virus infection in cynomolgus monkeys (*Macaca fascicularis*). *Lab. Invest.* **81**, 1581–1600 (2001).
- LeDuc, J. W., Damon, I., Relman, D. A., Huggins, J. & Jahrling, P. B. Smallpox research activities: U.S. interagency collaboration, 2001. *Emerg. Infect. Dis.* **8**, 743–745 (2002).
- Arita, I., Jezek, Z., Khodakevich, L. & Ruti, K. Human monkeypox: a newly emerged orthopoxvirus zoonosis in the tropical rain forests of Africa. *Am. J. Trop. Med. Hyg.* **34**, 781–789 (1985).
- Breman, J. G. & Henderson, D. A. Poxvirus dilemmas—monkeypox, smallpox, and biologic terrorism. *N. Engl. J. Med.* **339**, 556–559 (1998).

- Smith, G. L., Vanderplasm, A. & Law, M. The formation and function of extracellular enveloped vaccinia virus. *J. Gen. Virol.* **83**, 2915–2931 (2002).
- Law, M. & Smith, G. L. Antibody neutralization of the extracellular enveloped form of vaccinia virus. *Virology* **280**, 132–142 (2001).
- Hooper, J. W., Custer, D. M., Schmaljohn, C. S. & Schmaljohn, A. L. DNA vaccination with vaccinia virus L1R and A33R genes protects mice against a lethal poxvirus challenge. *Virology* **266**, 329–339 (2000).
- Hooper, J. W., Custer, D. M. & Thompson, E. Four-gene-combination DNA vaccine protects mice against a lethal vaccinia virus challenge and elicits appropriate antibody responses in nonhuman primates. *Virology* **306**, 181–195 (2003).
- Galmiche, M. C., Goenaga, J., Wittek, R. & Rindischbacher, L. Neutralizing and protective antibodies directed against vaccinia virus envelope antigens. *Virology* **254**, 71–80 (1999).
- Wolfe, E. J., Vijaya, S. & Moss, B. A myristylated membrane protein encoded by the vaccinia virus L1R open reading frame is the target of potent neutralizing monoclonal antibodies. *Virology* **211**, 53–63 (1995).
- Earl, P. L., Americo, J. L. & Moss, B. Development and use of a vaccinia virus neutralization assay based on flow cytometric detection of green fluorescent protein. *J. Virol.* **77**, 10684–10688 (2003).
- Appleyard, G., Hapel, A. J. & Boulter, E. A. An antigenic difference between intracellular and extracellular rabbitpox virus. *J. Gen. Virol.* **13**, 9–17 (1971).
- Law, M., Hollinshead, R. & Smith, G. L. Antibody-sensitive and antibody-resistant cell-to-cell spread by vaccinia virus: role of the A33R protein in antibody-resistant spread. *J. Gen. Virol.* **83**, 209–222 (2002).
- Breman, J. G. & Henderson, D. A. Diagnosis and management of smallpox. *N. Engl. J. Med.* **346**, 1300–1308 (2002).
- Speller, S. A. & Warren, A. P. *Ex vivo* detection and enumeration of human antigen-specific CD8⁺ T lymphocytes using antigen delivery by a recombinant vaccinia expression vector and intracellular cytokine staining. *J. Immunol. Methods* **262**, 167–180 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to R. Elkins (NIAID) for animal acquisition and to R. Byrum (Bioqual, Inc.) for animal care and immunizations. The work was supported by the National Institute of Allergy and Infectious Diseases.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.M. (bmoss@nih.gov).

Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression

Zhigao Wang¹, Da-Zhi Wang^{1,2}, Dirk Hockemeyer³, John McAnally¹, Alfred Nordheim³ & Eric N. Olson¹

¹Department of Molecular Biology, University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd, Dallas, Texas 75390-9148, USA

²Carolina Cardiovascular Biology Center, Department of Cell and Developmental Biology, University of North Carolina, Chapel Hill, North Carolina 27599-7126, USA

³Institute of Cell Biology, Department of Molecular Biology, Tuebingen University, D-72704, Tuebingen, Germany

Smooth muscle cells switch between differentiated and proliferative phenotypes in response to extracellular cues¹, but the transcriptional mechanisms that confer such phenotypic plasticity remain unclear. Serum response factor (SRF) activates genes involved in smooth muscle differentiation and proliferation by recruiting muscle-restricted cofactors, such as the transcriptional coactivator myocardin, and ternary complex factors (TCFs) of the ETS-domain family, respectively^{2–9}. Here we show that growth signals repress smooth muscle genes by triggering the displacement of myocardin from SRF by Elk-1, a TCF that acts as a myogenic repressor. The opposing influences of myocardin and Elk-1 on smooth muscle gene expression are mediated by structurally related SRF-binding motifs that compete for a common docking site on SRF. A mutant smooth muscle

promoter, retaining responsiveness to myocardin and SRF but defective in TCF binding, directs ectopic transcription in the embryonic heart, demonstrating a role for TCFs in suppression of smooth muscle gene expression *in vivo*. We conclude that growth and developmental signals modulate smooth muscle gene expression by regulating the association of SRF with antagonistic cofactors.

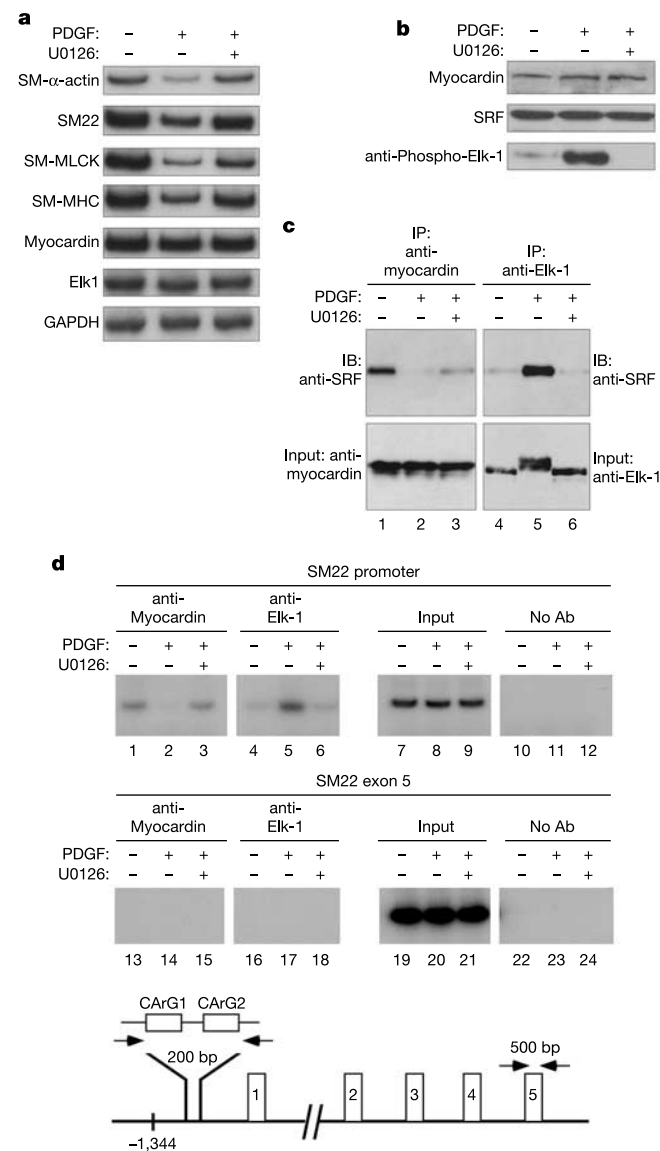


Figure 1 Mutually exclusive interactions of myocardin and Elk-1 with SRF in SM cells. A7r5 cells were cultured in serum-free medium and then stimulated with or without PDGF-BB (20 ng ml⁻¹) for 24 h (**a**) or 30 min (**b–d**) in the presence or absence of U0126 (10 μM) as indicated. **a**, RNA was isolated and transcripts detected by RT-PCR, using SM-α-actin, SM22, SM-myosin light chain kinase (MLCK) and SM-myosin heavy chain (MHC) as markers of smooth muscle differentiation and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a control for RNA loading. **b**, Extracts were prepared from A7r5 cells and proteins were detected by western blot. **c**, Immunoprecipitations (IP) performed in A7r5 cell extracts with anti-myocardin or anti-Elk-1 antibody followed by immunoblot (IB) with mouse anti-SRF antibody. **d**, Chromatin was prepared from A7r5 cells and immunoprecipitated with anti-myocardin or anti-Elk-1, followed by PCR using primers from the SM22 promoter or exon 5 of the SM22 gene as a control. ChIP assays with primers flanking exon 5 of the coding region showed no response to PDGF.

Most smooth muscle contractile protein genes are controlled by SRF¹⁰, which binds to a sequence known as a CARG box and recruits myocardin, a coactivator that is necessary and sufficient for smooth muscle gene expression^{3–8}. In response to growth factor signalling, SRF also interacts with TCFs, a family of ETS-domain transcription factors that includes Elk-1, SAP-1a and NET/SAP-2/Erp (refs 11–13). Phosphorylation of these TCFs by MAP kinases facilitates their association with SRF and activation of growth-factor-inducible genes^{14,15}.

To investigate how extracellular signals enable SRF to select between muscle-specific and growth-regulated genes, which represent opposing transcriptional programmes, we examined the effect of platelet-derived growth factor (PDGF), an antagonist of smooth muscle differentiation¹⁶, on myocardin/SRF target genes in the A7r5 smooth muscle (SM) cell line. PDGF specifically suppressed expression of contractile protein genes (Fig. 1a; Supplementary Fig. 1) and induced the phosphorylation of Elk-1, as detected by western blot with an anti-phospho-Elk-1 antibody or by an upshift of Elk-1 on SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 1b, c). PDGF also increased the association of Elk-1 with SRF, which was paralleled by a decrease in association of myocardin with SRF, as detected by co-immunoprecipitation (Fig. 1c).

The association of TCFs with SRF is stabilized on SRF-binding sites that contain adjacent ETS-binding sequences (GGAA/T) (ref. 17), which are present in several smooth muscle gene promoters. Chromatin immunoprecipitation (ChIP) assays showed that suppression of the differentiated phenotype by PDGF was accompanied by an increase in the association of Elk-1 with SRF-binding sites in the promoters of the SM22 and SM-α-actin genes, prototypical smooth muscle target genes of myocardin and SRF^{18–20} (lanes 4 and 5 in Fig. 1d; Supplementary Fig. 2). The PDGF-dependent recruitment of Elk-1 to these SRF-binding sites was accompanied by the disappearance of myocardin from the same sites (lanes 1 and 2 in Fig. 1d; Supplementary Fig. 2).

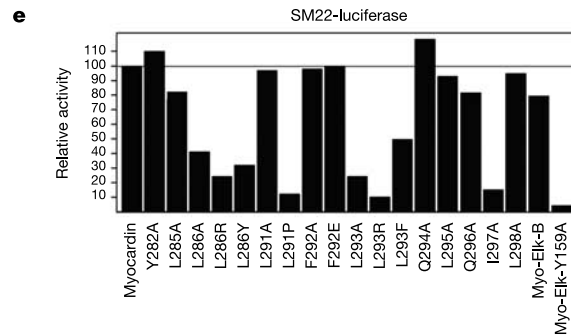
Inhibition by U0126 of the MAP kinases MEK-1/2 blocked the phosphorylation of Elk-1 and the suppressive effects of PDGF on smooth muscle gene expression (Fig. 1a; Supplementary Fig. 1). U0126 also prevented the signal-dependent association of Elk-1 with SRF and the displacement of myocardin (Fig. 1b–d). Thus, modulation by PDGF of the differentiated SM cell phenotype was accompanied by the exchange of Elk-1 for myocardin on SRF in the control regions of native myocardin target genes; MEK-1/2, which phosphorylates ERK-1/2, which in turn phosphorylates Elk-1 (refs 13, 21), was a key effector in this process.

Mutation of Val 194 to Glu in the DNA-binding domain of SRF abolishes association with TCFs²², and prevented ternary complex formation with myocardin (Fig. 2a), suggesting that these cofactors interact with the same region of SRF as has been observed for the myocardin-related transcription factor MAL²³. Indeed, co-incubation of Elk-1 and SRF with increasing amounts of myocardin resulted in a loss of the SRF–Elk-1 complex on the SRF-binding site within the SM22 promoter and the appearance of the SRF–myocardin complex (Fig. 2b, compare lanes 1–4 and 6–9). This competition required the basic region of myocardin (lane 10), which is essential for interaction with SRF (lane 5). Conversely, increasing the amount of Elk-1 in the DNA-binding assay resulted in the disappearance of the SRF–myocardin complex (Fig. 2b, compare lanes 11–14 and 16–19). Mutation of the TCF-binding site adjacent to the CARG box in the DNA probe did not affect the association of myocardin with SRF, but it impaired the ability of Elk-1 to compete with myocardin for interaction with SRF (Supplementary Fig. 3a).

TCFs interact with SRF through a B-box domain¹⁷. Replacement of Tyr 159 in the Elk-1 B-box with Ala abolishes interaction with SRF (lane 15 in Fig. 2b; ref. 24). This Elk-1 mutant (Y159A) was ineffective in displacing myocardin from SRF (lane 20 in Fig. 2b). Similar competition between myocardin and Elk-1 was observed

A myocardin mutant in which the 21-amino-acid B-box-like region was replaced with the B-box of Elk-1 (Myo-Elk-B) formed a

Myocardin was less effective in activating the SM22 promoter when cells were stimulated by 20% serum than by 0.5% serum. The reduction in its transcriptional activity in response to high serum



regions of TCFs and the corresponding region of myocardin and myocardin mutants are shown in pink shading. Residues in red are required for SRF interaction. The amino-terminal domain (NTD), basic region (++) and the glutamine-rich domain (Q) are shown.

d, Gel mobility shift assays were performed with SRF, myocardin, or the indicated myocardin mutants translated *in vitro* and a radiolabelled probe corresponding to the *c-fos* serum response element. The region of the gel containing shifted probe is shown.

e, COS cells were transiently transfected with expression vectors for myocardin or the indicated myocardin mutants and an *SM22*-luciferase reporter. Values are expressed as the level of luciferase activity with each mutant relative to that with myocardin, which was assigned a value of 100.

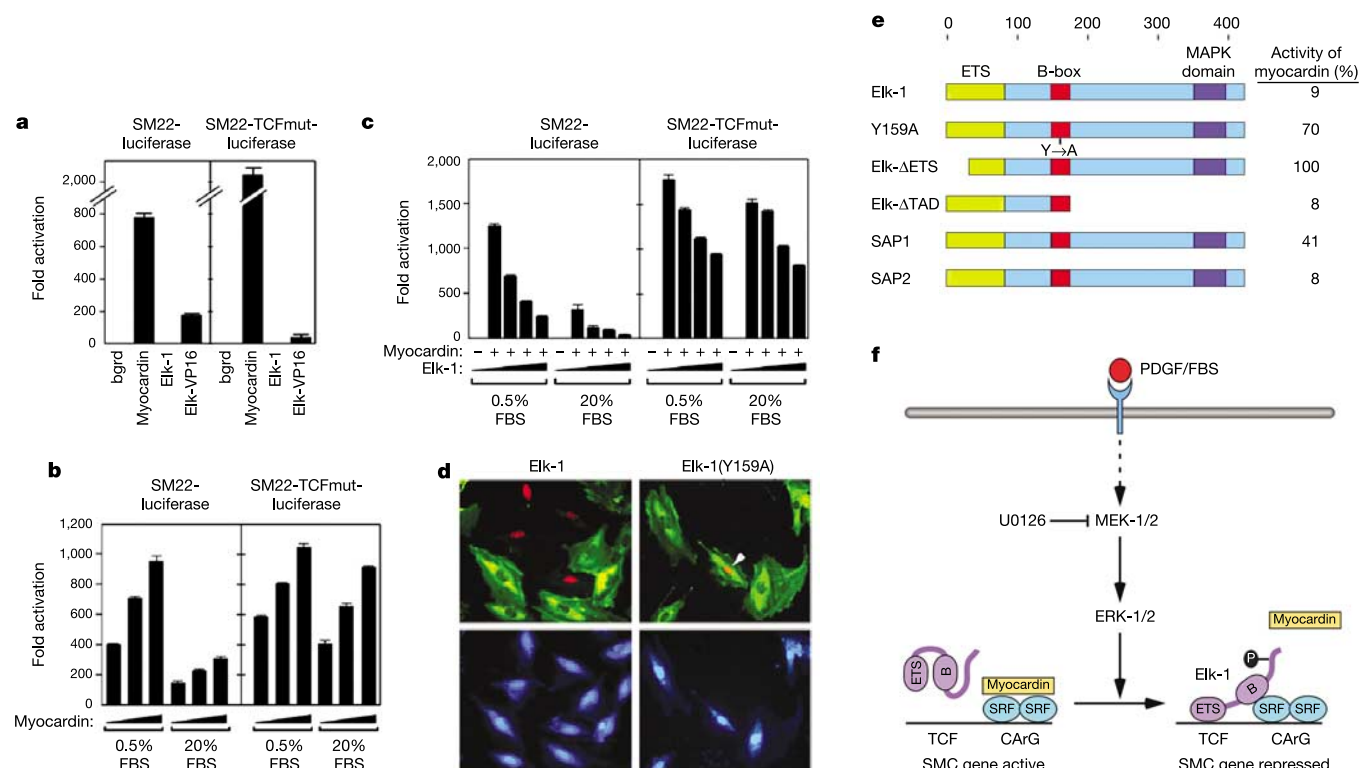


Figure 3 Suppression of smooth muscle gene expression by Elk-1. **a**, COS cells were transiently transfected with expression vectors for myocardin, Elk-1 or Elk-1-VP16 and luciferase reporters linked to the wild-type *SM22* promoter or the promoter with a mutation in the TCF site. Values are expressed as fold activation above background. **b**, COS cells were transfected with myocardin expression vector and the indicated reporters and maintained in 0.5% FBS or stimulated with 20% FBS for 6 h. **c**, COS cells were transfected with expression vectors encoding myocardin and increasing amounts of Elk-1 and the indicated reporters and cultured under the indicated conditions. **d**, A7r5 cells were transfected with expression plasmids encoding Elk-1 or Elk-1(Y159A) with Myc-epitope tags. Upper panels, cultures were stained for SM- α -actin (green) and Myc (red). Note that cells expressing Elk-1 fail to express

SM- α -actin (top left), whereas expression of Elk-1(Y159A) does not alter SM- α -actin expression (top right). Lower panels, Hoechst staining was used to detect nuclei. **e**, Mapping the regions of Elk-1 required for suppression of myocardin activity in transfected 10T1/2 cells. 10T1/2 cells were transfected with expression plasmids encoding myocardin and the indicated mutants of Elk-1. The relative number of SM-MHC-positive cells in each culture is expressed relative to the number in cultures transfected with myocardin alone, which was assigned a value of 100. Amino acids in Elk-1 are indicated along the top. **f**, A model to account for the modulation of smooth muscle genes by competition between myocardin and Elk-1 for SRF in response to growth signals.

was dependent on the TCF site (Fig. 3b) and was coupled to MEK-1/2-dependent phosphorylation of ERK-1/2, which stimulates the interaction of Elk-1 with SRF (Supplementary Fig. 5). Conversely, exogenous Elk-1 was a more effective repressor of the *SM22* promoter in the presence of 20% serum, and the TCF site enhanced its repressive activity (Fig. 3c). The Y159A mutant and an Elk-1 mutant lacking the ETS domain (Elk-1- Δ ETS), which mediates binding to the TCF sequence, were ineffective inhibitors of myocardin (Supplementary Fig. 6a). Elk-1 showed similar repressive activity against a reporter linked to three tandem copies of a CArG box and adjacent TCF site (Supplementary Fig. 6b), confirming that its effects are dependent on SRF rather than another regulator.

Elk-1, as well as SAP-1 and -2, suppressed activation of smooth muscle genes by myocardin in 10T1/2 cells and in the A7r5 smooth muscle cell line, whereas the Y159A mutant was ineffective (Fig. 3d, e; Supplementary Fig. 6c–e). Mutants of Elk-1 lacking either the ETS domain (Elk-1- Δ ETS) or a functional B-box were unable to interfere with myocardin activity (Fig. 3e). On the contrary, an Elk-1 mutant lacking the TAD was fully functional in suppressing myocardin activity (Fig. 3e, mutant Elk-1- Δ TAD).

To test whether TCFs modulate smooth muscle gene expression *in vivo*, we generated transgenic mouse embryos harbouring a *lacZ* transgene linked to the wild-type *SM22* promoter or the promoter

with a mutation in the TCF site. Both promoters directed high expression of *lacZ* in the developing vasculature (Fig. 4a). The wild-type *SM22* promoter and the endogenous *SM22* gene²⁶, like many smooth muscle genes, are also expressed in the heart until embryonic day (E)13.5. In contrast, the mutant *SM22* promoter was still expressed at high levels in the heart at E15.5, at least two days after cardiac expression of the wild-type promoter had been extinguished (Fig. 4a). Immunostaining of cardiac sections with an anti-phospho-Elk-1 antibody showed the presence of phospho-Elk-1 in cardiomyocytes at this stage, where it could serve to repress transcription of SRF-dependent smooth muscle genes (Fig. 4b). Thus, the TCF site is required for the precise temporal control of this smooth muscle promoter in the developing heart at a time when phospho-Elk-1 is expressed there.

Our results demonstrate that myocardin and Elk-1 compete for interaction with a common docking site on SRF and that Elk-1 acts as a signal-responsive repressor of smooth muscle gene expression by displacing myocardin from SRF within the context of native chromatin (Fig. 3f). The repressive influence of Elk-1 on smooth muscle genes is unexpected in the light of its role as an activator of growth-factor-inducible genes⁹. We propose that Elk-1, which is a weak activator compared with myocardin²⁷, is sufficient to suppress smooth muscle gene expression when bound to even a single SRF-binding site, owing to the dependence of myocardin on

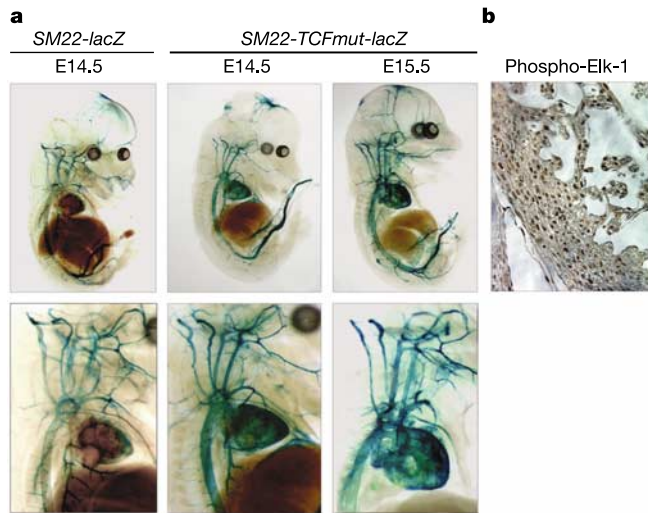


Figure 4 Ectopic expression of a *lacZ* reporter controlled by an *SM22* promoter lacking the TCF site. **a**, Transgenic mouse embryos of the indicated ages were stained for *LacZ* expression. The transgene with the mutation in the TCF site continues to be expressed in the heart after the wild-type promoter has been downregulated. **b**, Staining of a histological section of an E15.5 mouse heart with anti-phospho-Elk-1 antibody reveals the presence of phospho-Elk-1 in cardiomyocytes.

binding to multiple sites for maximal transcriptional activation^{3,28}. The mutually exclusive association of SRF with myocardin and Elk-1 (or other TCFs) provides a mechanism for modulating smooth muscle gene expression during development and in response to growth signalling. □

Methods

Cell culture and transfection

A7r5 cells were maintained as described previously¹⁶. Cells were treated with 20 ng ml⁻¹ PDGF-BB (Sigma) for 24 h and, when indicated, were pretreated with 10 μ M U0126 (Cell Signalling) for 2 h before PDGF-BB treatment. Transfections and luciferase assays were performed as described^{3,28}. Myocardin mutants were generated through polymerase chain reaction (PCR)-based mutagenesis using the QuickChange kit from Stratagene. The *SM22*-luciferase reporter contained the 1,434 base pair promoter¹⁸ and the mutant promoter contained a mutation in the TCF site.

ChIP and IP assays

ChIP assays were performed using the ChIP Assay Kit from Upstate Biotechnology as described in ref. 29. Immunoprecipitation assays were performed essentially as ChIP assays, except the immunoprecipitation products were separated on SDS-PAGE and analysed by western blot with anti-SRF antibody. Affinity-purified rabbit anti-myocardin antibody was generated against amino acids 141–159. Rabbit anti-Elk-1 antibodies were from Santa Cruz and Cell Signalling. Further details are in Supplementary Information.

Analysis of SM and 10T1/2 cells

Immunocytochemistry, myogenic assays, western blots, and reverse transcription (RT)-PCR were performed as described⁷.

Gel mobility shift assays

SRF, myocardin, and Elk-1 were translated *in vitro* with a TNT T7-coupled reticulocyte lysate system (Promega). Gel mobility shift assays were performed using double-stranded probes as described (ref. 3; Supplementary Information). The sequences of the distal *SM22*-CaRG box probes are shown below. The CaRG box is in bold and the TCF site is italicized. Wild-type: AGCTGTTTCAGGGTCCTGCCATAAAAGGTTTTCGCCG CCGCC; the TCF mutant: AGCTGTTTCAGGGTCCTGCCATAAAAGGTTTAA CCGCCGCC.

Histology

Generation of transgenic mice and *LacZ* staining were performed as described¹⁸. Paraffin sections of E15.5 mouse hearts were stained with anti-phospho-Elk-1 antibody according to manufacturer's instructions (Cell Signalling).

Received 17 December 2003; accepted 26 January 2004; doi:10.1038/nature02382.

1. Owens, G. K. Regulation of differentiation of vascular smooth muscle cells. *Physiol. Rev.* **75**, 487–517 (1995).

- Chang, D. F. *et al.* Cysteine-rich LIM-only proteins CRP1 and CRP2 are potent smooth muscle differentiation cofactors. *Dev. Cell* **4**, 107–118 (2003).
- Wang, D. *et al.* Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. *Cell* **105**, 851–862 (2001).
- Du, K. L. *et al.* Myocardin is a critical serum response factor cofactor in the transcriptional program regulating smooth muscle cell differentiation. *Mol. Cell. Biol.* **23**, 2425–2437 (2003).
- Chen, J., Kitchen, C. M., Streb, J. W. & Miano, J. M. Myocardin: a component of a molecular switch for smooth muscle differentiation. *J. Mol. Cell. Cardiol.* **34**, 1345–1356 (2002).
- Yoshida, T. *et al.* Myocardin is a key regulator of CaRG-dependent transcription of multiple smooth muscle marker genes. *Circ. Res.* **92**, 856–864 (2003).
- Wang, Z., Wang, D. Z., Pipes, G. C. & Olson, E. N. Myocardin is a master regulator of smooth muscle gene expression. *Proc. Natl Acad. Sci. USA* **100**, 7129–7134 (2003).
- Li, S., Wang, D. Z., Wang, Z., Richardson, J. A. & Olson, E. N. The serum response factor coactivator myocardin is required for vascular smooth muscle development. *Proc. Natl Acad. Sci. USA* **100**, 9366–9370 (2003).
- Treisman, R. Ternary complex factors: growth factor regulated transcriptional activators. *Curr. Opin. Genet. Dev.* **4**, 96–101 (1994).
- Miano, J. Serum response factor: toggling between disparate programs of gene expression. *J. Mol. Cell. Cardiol.* **35**, 577–593 (2003).
- Shaw, P. E., Schroter, H. & Nordheim, A. The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. *Cell* **56**, 563–572 (1989).
- Giovane, A., Pintzas, A., Maira, S. M., Sobieszczuk, P. & Wasylyk, B. Net, a new ets transcription factor that is activated by Ras. *Genes Dev.* **8**, 1502–1513 (1994).
- Sharrocks, A. D. Complexities in ETS-domain transcription factor function and regulation: lessons from the TCF (ternary complex factor) subfamily. The Colworth Medal Lecture. *Biochem. Soc. Trans.* **30**, 1–9 (2002).
- Janknecht, R., Ernst, W. H., Pingoud, V. & Nordheim, A. Activation of ternary complex factor Elk-1 by MAP kinases. *EMBO J.* **12**, 5097–5104 (1993).
- Marais, R., Wynne, J. & Treisman, R. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell* **73**, 381–393 (1993).
- Holycross, B. J., Blank, R. S., Thompson, M. M., Peach, M. J. & Owens, G. K. Platelet-derived growth factor-BB-induced suppression of smooth muscle cell differentiation. *Circ. Res.* **71**, 1525–1532 (1992).
- Shore, P. *et al.* Determinants of DNA-binding specificity of ETS-domain transcription factors. *Mol. Cell. Biol.* **16**, 3338–3349 (1996).
- Li, L., Liu, Z., Mercer, B., Overbeck, P. & Olson, E. N. Evidence for serum response factor-mediated regulatory networks governing *SM22* transcription in smooth, skeletal, and cardiac muscle cells. *Dev. Biol.* **187**, 311–321 (1997).
- Kim, S., Ip, H. S., Lu, M. M., Clendenen, C. & Parmacek, M. S. A serum response factor-dependent transcriptional regulatory program identifies distinct smooth muscle cell sublineages. *Mol. Cell. Biol.* **17**, 2266–2278 (1997).
- Mack, C. P. & Owens, G. K. Regulation of smooth muscle α -actin expression *in vivo* is dependent on CaRG elements within the 5' and first intron promoter regions. *Circ. Res.* **84**, 852–861 (1999).
- Heldin, C. H., Ostman, A. & Ronnstrand, L. Signal transduction via platelet-derived growth factor receptors. *Biochim. Biophys. Acta* **1378**, F79–113 (1998).
- Ling, Y., West, A. G., Roberts, E. C., Lakey, J. H. & Sharrocks, A. D. Interaction of transcription factors with serum response factor. Identification of the Elk-1 binding surface. *J. Biol. Chem.* **273**, 10506–10514 (1998).
- Miralles, F., Posern, G., Zaromytidou, A. I. & Treisman, R. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113**, 329–342 (2003).
- Ling, Y., Lakey, J. H., Roberts, E. C. & Sharrocks, A. D. Molecular characterization of the B-box protein-protein interaction motif of the ETS-domain transcription factor Elk-1. *EMBO J.* **16**, 2431–2440 (1997).
- Hassler, M. & Richmond, T. J. The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *EMBO J.* **20**, 3018–3028 (2001).
- Li, L., Miano, J. M., Cserjesi, P. & Olson, E. N. *SM22*, a marker of adult smooth muscle, is expressed in multiple myogenic lineages during embryogenesis. *Circ. Res.* **78**, 188–195 (1996).
- Cen, B. *et al.* Megakaryoblastic leukemia 1, a potent transcriptional activator for serum response factor (SRF), is required for serum induction of SRF target genes. *Mol. Cell. Biol.* **23**, 6597–6608 (2003).
- Wang, D. Z. *et al.* Potentiation of serum response factor activity by a family of myocardin-related transcription factors. *Proc. Natl Acad. Sci. USA* **99**, 14855–14860 (2002).
- Qiu, P. & Li, L. Histone acetylation and recruitment of serum responsive factor and CREB-binding protein onto *SM22* promoter during *SM22* gene expression. *Circ. Res.* **90**, 858–865 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Sharrocks, M. Parmacek and L. Schuger for reagents, A. Tizenor for graphics, J. Page for editorial assistance, L. Sutherland for technical assistance and R. Bassel-Duby, L. Li, M. Tallquist and H. Yanagisawa for discussions. This work was supported by grants from the NIH, the McGowan Foundation, the Donald W. Reynolds Foundation, and the Robert A. Welch Foundation to E.N.O., from the Muscular Dystrophy Association (MDA) and NIH to D.-Z.W., and by the DFG to A.N.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.N.O. (Eric.Olson@utsouthwestern.edu).

Control of the SCF^{Skp2-Cks1} ubiquitin ligase by the APC/C^{Cdh1} ubiquitin ligase

Tarig Bashir, N. Valerio Dorrello, Virginia Amador, Daniele Guardavaccaro & Michele Pagano

Department of Pathology, MSB 599, New York University School of Medicine, and New York University Cancer Institute, 550 First Avenue, New York, New York 10016, USA

Skp2 and its cofactor Cks1 are the substrate-targeting subunits of the SCF^{Skp2-Cks1} (Skp1/Cul1/F-box protein) ubiquitin ligase complex that regulates entry into S phase by inducing the degradation of the cyclin-dependent kinase inhibitors p21 and p27 (ref. 1). Skp2 is an oncoprotein that often shows increased expression in human cancers²; however, the mechanism that regulates its cellular abundance is not well understood. Here we show that both Skp2 and Cks1 proteins are unstable in G1 and that their degradation is mediated by the ubiquitin ligase APC/C^{Cdh1} (anaphase-promoting complex/cyclosome and its activator Cdh1). Silencing of Cdh1 by RNA interference in G1 cells stabilizes Skp2 and Cks1, with a consequent increase in p21 and p27 proteolysis. Depletion of Cdh1 also increases the percentage of cells in S phase, whereas concomitant downregulation of Skp2 reverses this effect, showing that Skp2 is an essential

target of APC/C^{Cdh1}. Expression of a stable Skp2 mutant that cannot bind APC/C^{Cdh1} induces premature entry into S phase. Thus, the induction of Skp2 and Cks1 degradation in G1 represents a principal mechanism by which APC/C^{Cdh1} prevents the unscheduled degradation of SCF^{Skp2-Cks1} substrates and maintains the G1 state.

We examined the cell-cycle regulation of Skp2 and Cks1 proteins and found that they decreased considerably in late M phase and G1, and were induced again as cells approached S phase (Fig. 1a). To assess whether the reduction in Skp2 and Cks1 was due to protein destabilization, we measured their half-lives in prometaphase cells and in cells released from a prometaphase block for 1 h. We found that both Skp2 and Cks1 are more stable in prometaphase than in late M/early G1 (Fig. 1b). Furthermore, in cells exiting the cell cycle, Cks1 was found to decrease in parallel with Skp2 (Fig. 1c), which has been previously reported to decrease in response to serum deprivation³. Treatment with the proteasome inhibitor ZLLL prevented the downregulation of both Skp2 and Cks1 (Fig. 1d). Thus, we conclude that Skp2 and Cks1 proteins are unstable in G0/G1.

The oscillations of Skp2 and Cks1 roughly paralleled those of cyclin A and cyclin B (Fig. 1a), which suggested that they might be degraded through a similar mechanism. We therefore investigated the involvement of the ubiquitin ligase APC/C⁴ in controlling the degradation of Skp2 and Cks1. Because APC/C is regulated by its association with one of two activator proteins, Cdh1 and Cdc20, we transfected 293T cells with a construct expressing either Skp2 (Fig. 1e) or Cks1 (Fig. 1f), along with a construct expressing Cdh1 or Cdc20. Enforced expression of Cdh1 resulted in a considerable

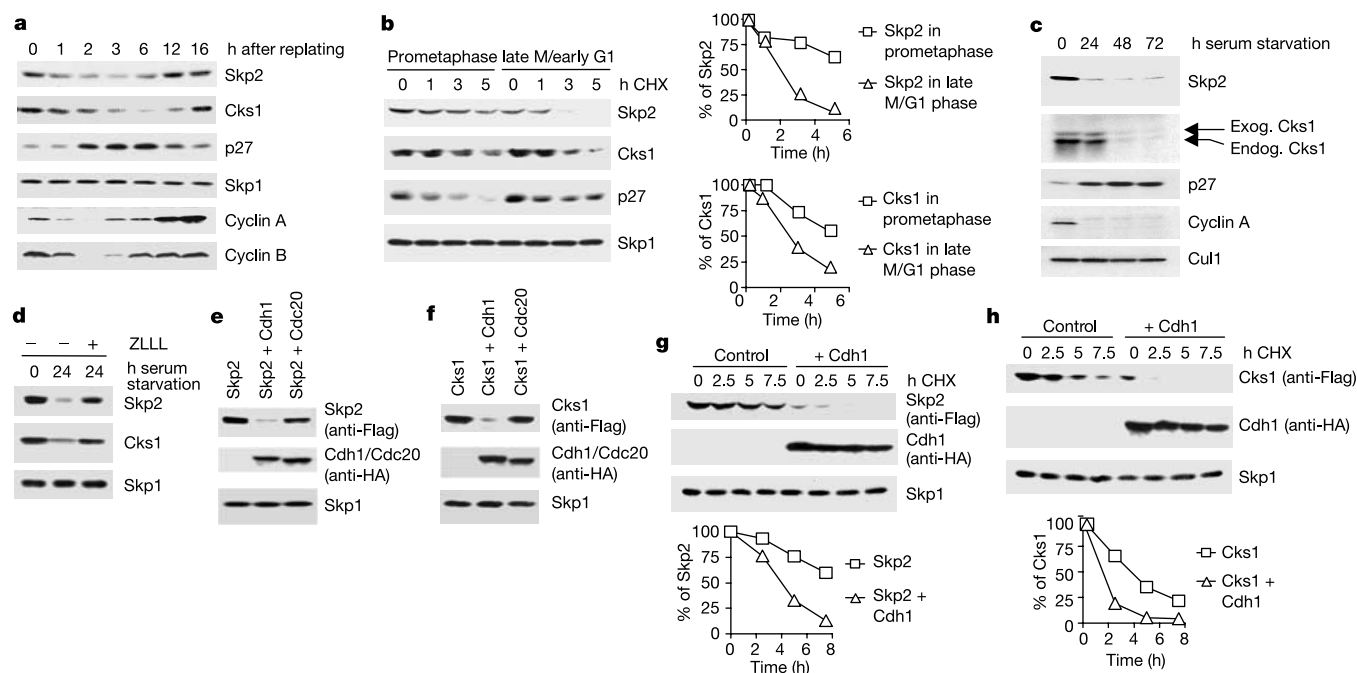


Figure 1 *In vivo* degradation of Skp2 and Cks1 is stimulated by Cdh1. **a–h**, Immunoblots. **a**, HeLa cells were released from nocodazole-induced prometaphase arrest and collected at the indicated times. **b**, Prometaphase cells were replated for 1 h in the presence (prometaphase) or absence (late M/early G1) of nocodazole. Cycloheximide (CHX) was then added and cell extracts were prepared at the indicated times. The results are quantified in the graphs on the right. **c**, Rat-1 cells infected with a retrovirus expressing Flag-tagged Cks1 were serum-deprived and extracts were prepared at the indicated times. **d**, Rat-1 cells were serum-deprived for 24 h and the proteasome inhibitor ZLLL (+) or vehicle alone (–) was added for the last 6 h. **e**, Coexpression of Flag-Skp2 and either HA-Cdh1 or HA-Cdc20 in 293T cells. **f**, Coexpression of Flag-Cks1 and either HA-Cdh1 or HA-Cdc20 in 293T cells. **g**, HeLa cells were transfected with Flag-Skp2 with or without HA-Cdh1. Twenty-four hours after transfection, cells were treated with CHX. Skp2 expression is quantified in the graph below. **h**, HeLa cells were transfected with Flag-Cks1 with or without HA-Cdh1. The half-life of Cks1 was analysed as in **g**.

or vehicle alone (–) was added for the last 6 h. **e**, Coexpression of Flag-Skp2 and either HA-Cdh1 or HA-Cdc20 in 293T cells. **f**, Coexpression of Flag-Cks1 and either HA-Cdh1 or HA-Cdc20 in 293T cells. **g**, HeLa cells were transfected with Flag-Skp2 with or without HA-Cdh1. Twenty-four hours after transfection, cells were treated with CHX. Skp2 expression is quantified in the graph below. **h**, HeLa cells were transfected with Flag-Cks1 with or without HA-Cdh1. The half-life of Cks1 was analysed as in **g**.

decrease in the steady-state amounts of both proteins, whereas overexpression of Cdc20 had no effect. Notably, the decrease in Skp2 and Cks1 expression correlated with their destabilization (Fig. 1g, h).

We next used small interfering RNAs (siRNAs) to reduce the expression of Cdh1. After transfection with siRNA, prometaphase HeLa cells were collected and replated in fresh medium. Cells transfected with siRNA targeting Cdh1 showed no G1-specific reduction in Skp2 and Cks1, a much less pronounced upregulation of p21 and p27 as compared with control cells (Fig. 2a), and an increase in the half-life of both Skp2 and Cks1 (Fig. 2b).

We also silenced Cdh1 expression in T98G cells exiting the cell cycle (as shown by the accumulation of p27; Fig. 2c). Silencing of Cdh1 prevented the degradation of Skp2 and Cks1 for up to 36 h after serum removal, when the effect of the siRNA started to wane, as judged by the recovery of Cdh1. It has been suggested that Skp2 downregulation in G0/G1 is due to Cul1-dependent auto-ubiquitination³; however, depletion of Cul1 did not prevent the degradation of Skp2 and Cks1 (Fig. 2c).

Downregulation of Cdh1 resulted in an increase in the number of HeLa cells incorporating 5-bromodeoxyuridine (BrdU; Fig. 2d). Notably, co-depletion of Skp2 reversed the stimulatory effect

induced by Cdh1 inhibition. Furthermore, single-cell immunofluorescence analysis (Fig. 2d, bottom) showed that Cdh1-depleted cells progressed faster through S phase, as $78.2 \pm 1.4\%$ of these cells showed a diffuse BrdU staining typical of advanced S phase, rather than the distinct foci of BrdU incorporation characteristic of initial DNA replication. By contrast, diffuse BrdU staining was present in only $65.7 \pm 2.1\%$ of control cells, $37.8 \pm 0.5\%$ of Skp2-depleted cells and $35 \pm 2.1\%$ of cells depleted in both Cdh1 and Skp2. Taken together, these results show that in G0/G1 the stability of Skp2 and Cks1 is regulated by APC/C^{Cdh1} and that Skp2 is an essential target of Cdh1, because in Skp2-depleted cells Cdh1 downregulation is insufficient to promote DNA replication.

Additional key observations could be derived from the siRNA experiments with regard to the cell cycle in the presence of reduced amounts of Cdh1. First, cells still exited mitosis with normal kinetics, as judged by the degradation of cyclin A and cyclin B and the disappearance of phosphorylation on Ser 10 of histone H3 (Fig. 2a). By contrast, silencing of Cdc20 delayed mitotic exit by about 1 h (data not shown). Second, cyclin A and cyclin B are normally downregulated in mitosis, but cyclin A re-accumulated at G1/S slightly earlier in Cdh1-silenced cells than in control cells (Fig. 2a, time points 4 and 6). This effect was even more evident in

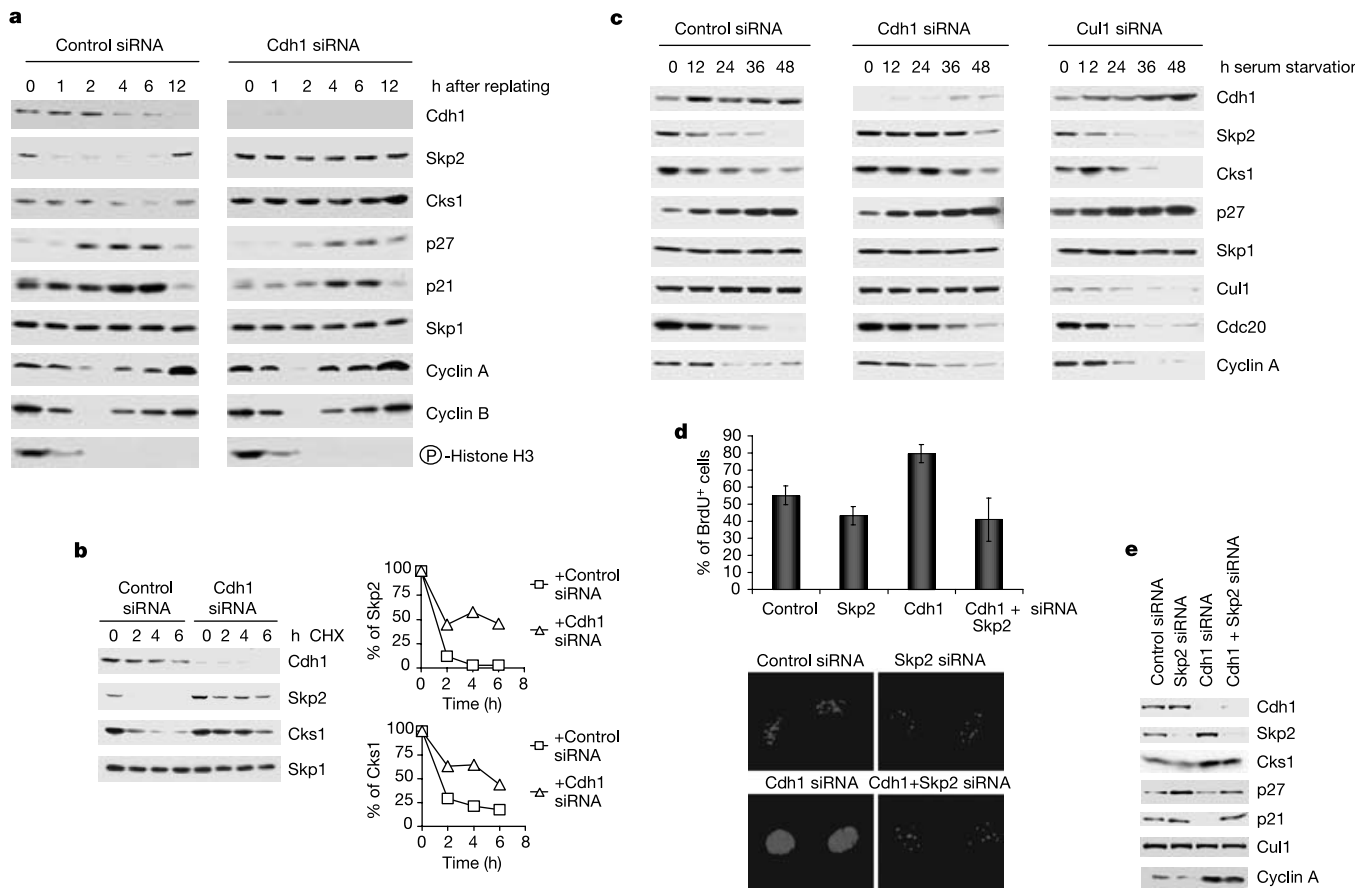


Figure 2 Cdh1 silencing stabilizes Skp2 and Cks1 in G1 and G0. **a–c**, Immunoblots. **a**, HeLa cells were transfected with siRNA corresponding to a non-relevant mRNA (control siRNA) or to Cdh1 mRNA. Twelve hours after the last transfection, nocodazole was added for 16 h. Prometaphase cells were collected and replated in fresh medium for the indicated times. **b**, Silencing of Cdh1 and synchronization were done as in **a**. Prometaphase cells were replated in drug-free medium for 1 h before CHX was added for the indicated times. The results are quantified in the graphs on the right. **c**, Silencing of

Cdh1 and Cul1 was done as in **a**, except that T98G cells were used. Twenty-four hours after the last transfection, cells were deprived of serum for the indicated times. **d**, Top, HeLa cells were transfected with the indicated siRNAs and labelled with BrdU for 60 min, and the percentage of BrdU-positive cells was determined by immunofluorescence; bottom, representative cells visualized by immunofluorescence using an antibody against BrdU. **e**, Cells treated as in **d** were analysed by immunoblotting.

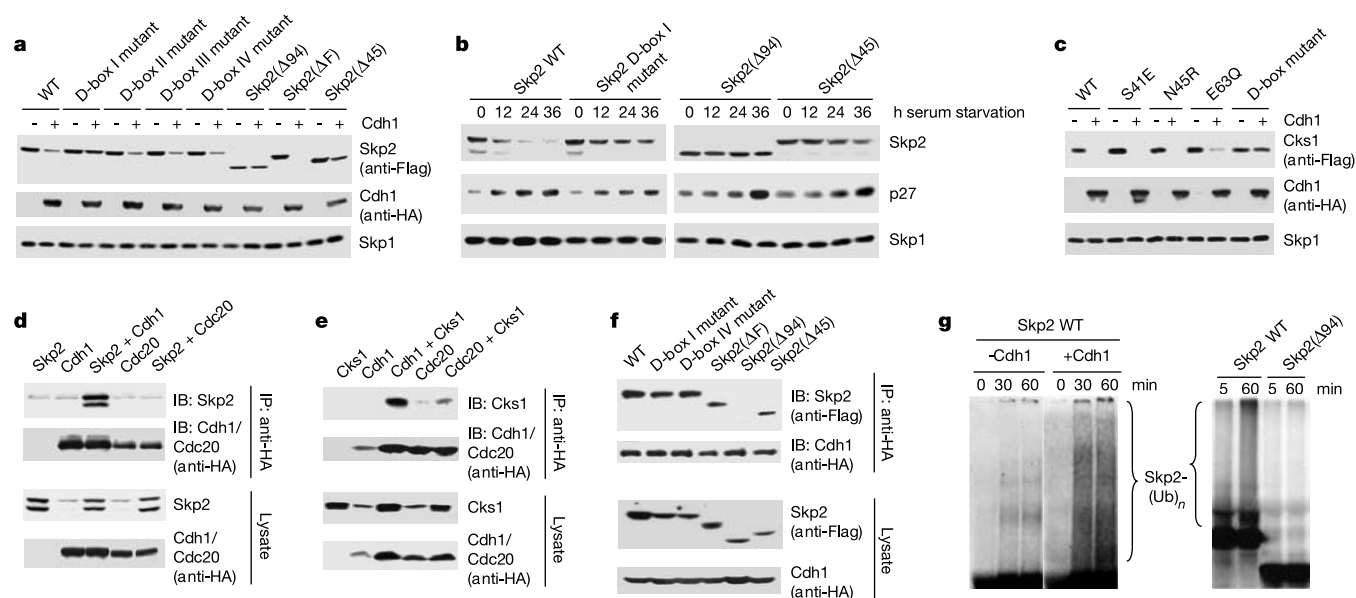


Figure 3 Requirements for APC/C^{Cdh1}-mediated degradation of Skp2 and Cks1. **a–f**, Immunoblots. **a**, Expression of Flag-tagged wild-type Skp2 or the indicated Skp2 mutants with or without HA-Cdh1 in 293T cells. **b**, Rat-1 cells were infected with retroviruses expressing human wild-type Skp2 or the indicated Skp2 mutants, and deprived of serum for the indicated times. **c**, Expression of Flag-tagged wild-type Cks1 or the indicated Cks1 mutants with or without HA-Cdh1 in 293T cells. **d**, Skp2 and either HA-Cdh1 or HA-Cdc20 were expressed in 293T cells, and ZLL was added 6 h before the

cells were collected. Cell extracts were immunoprecipitated (IP) with an antibody against HA and analysed by immunoblotting (IB). **e**, As **d**, except that Cks1 rather than Skp2 was expressed. **f**, As **d**, but with the indicated Skp2 mutants. **g**, Immunopurified APC/C supplemented with recombinant Cdh1 ubiquitinates *in vitro* translated wild-type Skp2 but not Skp2(Δ 94). The brackets mark a ladder of bands of relative molecular mass >45,000 corresponding to polyubiquitinated Skp2.

asynchronous cells, in which the silencing of Cdh1 induced a marked increase in cyclin A (Fig. 2e). Third, the accumulation of p27 and p21 was markedly reduced in cells re-entering G1 from mitosis (Fig. 2a) and in asynchronous cells (Fig. 2e), but not in serum-starved cells (Fig. 2c and data not shown). Thus, Cdh1 seems to be rate limiting for the accumulation of SCF^{Skp2-Cks1} substrates in cycling cells, but not in cells withdrawing from the cell cycle.

Substrates of APC/C are often characterized by the presence of a 'destruction box' (D-box), which was first identified in cyclin B as a stretch of nine amino acids (RxxLxxIxN) that when mutated stabilizes the protein⁴. However, many substrates contain only the minimal RxxL motif in their D-boxes, which together with additional regions in the substrate regulate their degradation by APC/C. We identified five RxxL motifs in human Skp2 that potentially might be D-boxes (amino acids 3–6, 84–87, 234–237, 294–297 and 415–418), but no other known putative Cdh1 recognition motifs (such as the KEN-box or A-box). One potential D-box (amino acids 234–237) is located in a leucine-rich repeat, which is part of the putative substrate-binding motif and therefore less likely to be also a recognition site for Cdh1.

We mutated the arginine and leucine residues in the other four motifs to alanine and determined the stabilities of these D-box mutants and three Skp2 deletion mutants in 293T cells overexpressing Cdh1. Of these Skp2 mutants, one with a mutation at the very amino terminus (D-box I) and two N-terminal deletion mutants, Skp2(Δ 45) and Skp2(Δ 94), were resistant to the effect of overexpressed Cdh1 (Fig. 3a). Notably, a Skp2 mutant lacking the F-box, Skp2(Δ F), was efficiently downregulated by Cdh1, showing that Skp2 does not need to be part of an SCF complex to be a substrate of APC/C. We also investigated the stabilities of the D-box I and deletion mutants in Rat-1 fibroblasts exiting the cell cycle. Skp2 was monitored by immunoblotting with an antibody against human Skp2 that does not recognize endogenous rat Skp2

(ref. 5). All three Skp2 mutants were clearly stabilized under conditions that promoted rapid degradation of wild-type Skp2 (Fig. 3b), confirming that the degradation of Skp2 is dependent on D-box I.

We next investigated the requirements for Cdh1-mediated degradation of Cks1. There is one RxxL motif at the carboxy terminus of Cks1 (amino acids 70–73), which when mutated prevented Cdh1-induced degradation (Fig. 3c), suggesting that this region represents a bona fide D-box. In addition, a Cks1 mutant that cannot bind cyclin-dependent kinases (CDKs), Glu63Gln, and two mutants that cannot bind Skp2, Ser41Glu and Asn45Arg (ref. 6), were still downregulated when Cdh1 was coexpressed (Fig. 3c), showing that stable binding to either CDKs or Skp2 is not a prerequisite for APC/C^{Cdh1}-mediated degradation of Cks1.

Furthermore, we found that Skp2 physically interacts with Cdh1 but not with Cdc20 (Fig. 3d), confirming the specificity of Cdh1 for this substrate. Similarly, Cks1 co-immunoprecipitated with Cdh1 and to a lesser extent with Cdc20 (Fig. 3e), which is not surprising because Cks1 binds to APC/C^{Cdc20} to promote its phosphorylation⁷. Notably, wild-type Skp2, but not Skp2(Δ 94), was detected in anti-Cdh1 immunoprecipitates (Fig. 3f). Although the Skp2 D-box I mutant and Skp2(Δ 45) were more stable than the wild-type protein (Fig. 3a, b), they still bound Cdh1 (Fig. 3f). These results show that a region of Skp2 between amino acids 46 and 94 mediates the binding between Skp2 and Cdh1, in line with the notion that the RxxL motif is an APC/C recognition motif but does not necessarily mediate a stable interaction with this ligase. The importance of the Skp2 N-terminal region was further shown by the ability of APC/C^{Cdh1} to ubiquitinate *in vitro* wild-type Skp2 but not Skp2(Δ 94) (Fig. 3g). In contrast to Skp2, we were unable to *in vitro* ubiquitinate Cks1 using APC/C^{Cdh1}, suggesting that an essential factor was missing or that Cks1, although downstream of APC/C^{Cdh1}, might not be its direct substrate.

To examine further the biological significance of Skp2 destruc-

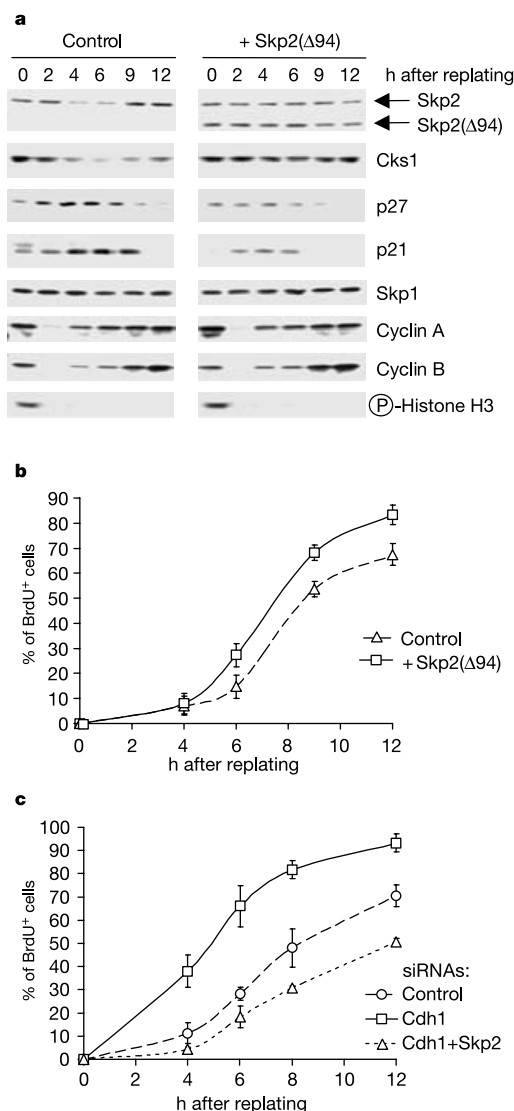


Figure 4 Biological relevance of Skp2 degradation. **a**, HeLa cells were either mock infected or infected with a retrovirus expressing Skp2(Δ94). After nocodazole treatment, prometaphase cells were replated in fresh medium for the indicated times. Cells were collected and analysed by immunoblotting. **b**, The percentage of BrdU-positive cells was determined in three independent experiments done as in **a**. **c**, The percentage of BrdU-positive cells was determined in three independent experiments done as in Fig. 2d, with the exception that cells were synchronized as in Fig. 2a and released from prometaphase for the indicated times before BrdU was added for 30 min.

tion, we evaluated the effect of expressing the stable Skp2(Δ94) mutant in synchronized cells. As compared with control cells, cells expressing Skp2(Δ94) showed accelerated degradation of p21 and p27, and stabilization of endogenous Skp2 and Cks1 in G1 (Fig. 4a). Notably, these cells progressed considerably faster into S phase (Fig. 4b), although not to the same extent as Cdh1-depleted cells (Fig. 4c). This difference is probably due to the fact that Skp2 is an essential, but not the only, substrate that is degraded through APC/C^{Cdh1} to prevent premature entry into S phase. Extensive overexpression of Skp2 achieved with adenoviral vectors accelerates S-phase entry⁸, whereas physiological amounts of Skp2 expressed from retroviral vectors are subject to regulated proteolysis and

are unable to force cells into S phase (refs 3, 5, and Fig. 3b). By contrast, the expression of physiological amounts of a stable Skp2 mutant in G1 induced premature S-phase entry. These results show that Skp2 destruction is an essential event for proper cell-cycle control.

Experiments in cultured cells and transgenic mice have shown that Skp2 is the product of a proto-oncogene². In human carcinomas and lymphomas, p27 destabilization is predictive of poor prognosis and correlates with high concentrations of Skp2 (ref. 9). We have shown here that the degradation of Skp2 and Cks1 is controlled by APC/C^{Cdh1}, which represents a principal mechanism to maintain the G1 state and to avoid premature entry into S phase, a potential cause of genetic instability. Deregulation of this mechanism might contribute to overexpression of Skp2 and Cks1 and thereby promote uncontrolled proliferation in cancer cells. □

Methods

Cell culture and biochemistry

Cell culture and synchronization of HeLa, Rat-1, 293T and T98G cells were done as described^{3,5,10}. Antibodies against Skp2, Skp1, p27, p21, cyclin A, cyclin B, Cull1, phospho-histone H3, BrdU, haemagglutinin (HA) and Flag have been described^{5,10}. Antibodies against Cdh1 and Cdc20 were a gift from J. Lukas. An antibody against human Cks1 (which does not crossreact with Cks2) was generated in collaboration with Zymed. Immunoblot analysis, immunoprecipitations and immunofluorescence were done as described¹⁰. We carried out *in vitro* ubiquitination as described¹¹, except that we used extracts from serum-deprived T98G cells.

Plasmids, transfections, infections and siRNAs

Skp2 and Cks1 mutants were generated using the QuickChange Site-directed Mutagenesis kit (Stratagene). Cell transfections and retroviral infections were done as described^{10,12}. The siRNA oligonucleotide sequences were 5'-AATGAGAAGTCT CCCAGTCAGdTdT-3' (Cdh1), 5'-AAGGAGUGACAAGACUUGdTdT-3' or 5'-AAUCUAAAGCCUGGAA GGCCUGdTdT-3' (Skp2) and 5'-UAGACAUUGGGUUCGCCUGdTdT-3' (Cul1).

Received 24 November 2003; accepted 9 January 2004; doi:10.1038/nature02330.

- Reed, S. Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nature Rev. Mol. Cell Biol.* **4**, 855–864 (2003).
- Pagano, M. & Benmaamar, R. When protein destruction runs amok, malignancy is on the loose. *Cancer Cell* **4**, 251–256 (2003).
- Wirtelbauer, C. *et al.* The F-box protein Skp2 is a ubiquitylation target of a Cul1-based core ubiquitin ligase complex: evidence for a role of Cul1 in the suppression of Skp2 expression in quiescent fibroblasts. *EMBO J.* **19**, 5362–5375 (2000).
- Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
- Carrano, A. C. & Pagano, M. Role of the F-box protein Skp2 in adhesion-dependent cell cycle progression. *J. Cell Biol.* **153**, 1381–1389 (2001).
- Sitry, D. *et al.* Three different binding sites of Cks1 are required for p27–ubiquitin ligation. *J. Biol. Chem.* **277**, 42233–42240 (2002).
- Sudakin, V., Shtenberg, M., Ganoh, D., Hershko, J. & Hershko, A. Binding of activated cyclosome to p13^{Suc1}. Use for affinity purification. *J. Biol. Chem.* **272**, 18051–18059 (1997).
- Sutterluty, H. *et al.* p45^{Skp2} promotes p27^{Kip1} degradation and induces S phase in quiescent cells. *Nature Cell Biol.* **1**, 207–214 (1999).
- Bloom, J. & Pagano, M. Deregulated degradation of the CDK inhibitor p27 and malignant transformation. *Semin. Cancer Biol.* **13**, 41–47 (2003).
- Guardavaccaro, D. *et al.* Control of meiotic and mitotic progression by the F box protein β-Trcp1 *in vivo*. *Dev. Cell* **4**, 799–812 (2003).
- Fang, G., Yu, H. & Kirschner, M. W. Direct binding of Cdc20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Mol. Cell* **2**, 163–171 (1998).
- Bloom, J., Amador, V., Bartolini, F., DeMartino, G. & Pagano, M. Proteasome-mediated degradation of p21 via N-terminal ubiquitylation. *Cell* **115**, 71–82 (2003).

Acknowledgements We thank A. Hershko for contributing to this work; W. Kaelin and M. Kirschner for communicating results before publication; J. Bloom, T. Cardozo, L. Gardner, A. Hershko and E. Oxelmark for critically reading the manuscript; and M. Brandeis, B. Dynlacht, G. Draetta, K. Helin and J. Lukas for reagents. M.P. is grateful to T. M. Thor for continuous support. This work was supported by a fellowship from the New York State Breast Cancer Research and Education fund to T.B.; a Federazione Italiana Ricerca sul Cancro fellowship to N.V.D.; a Human Frontiers Science Program fellowship to V.A.; an Italian American Cancer Foundation fellowship and a Susan Komen Breast Cancer Foundation fellowship to D.G.; and a Irma Hirsch Scholarship and NIH grants to M.P.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P. (michele.pagano@med.nyu.edu).

Degradation of the SCF component Skp2 in cell-cycle phase G1 by the anaphase-promoting complex

Wenyi Wei^{1*}, Nagi G. Ayad^{2*}, Yong Wan², Guo-Jun Zhang¹, Marc W. Kirschner² & William G. Kaelin Jr^{1,3}

¹Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, ²Department of Cell Biology, and ³Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

Cell-cycle transitions are driven by waves of ubiquitin-dependent degradation of key cell-cycle regulators. SCF (Skp1/Cullin/F-box protein) complexes and anaphase-promoting complexes (APC) represent two major classes of ubiquitin ligases whose activities are thought to regulate primarily the G1/S and metaphase/anaphase cell-cycle transitions, respectively^{1,2}. The major target of the Skp1/Cul1/Skp2 (SCF^{SKP2}) complex is thought to be the Cdk inhibitor p27 during S phase, whereas the principal targets for the APC are thought to be involved in chromatid separation (securin) and exit from mitosis (cyclin B). Although the role of the APC in mitosis is relatively clear, there is mounting evidence that APCs containing Cdh1 (APC^{CDH1}) also have a function in the G1 phase of the cell cycle^{2,3}. Here, we show that the F-box protein Skp2 is polyubiquitinated, and hence earmarked for destruction, by APC^{CDH1}. As a result, accumulation of SCF^{SKP2} requires prior inactivation of APC^{CDH1}. These findings provide an insight into the orchestration of SCF and APC activities during cell-cycle progression, and into the involvement of the APC in G1.

Both transcriptional and post-transcriptional changes are thought to underlie cell-cycle-dependent changes in Skp2 (refs 4–6). Skp2

levels are lowest during the G1 phase of the cell cycle, a period when the APC^{CDH1} ubiquitin ligase complex is active. We noted that Skp2 contains a canonical D-box motif (R-X-X-L-X-X-X-N/D/E)⁷, suggesting that it might be a direct substrate of APC^{CDH1} (Fig. 1a). In support of these clues, we found evidence of Skp2 bound to Cdh1 in co-immunoprecipitation and glutathione S-transferase (GST) pull-down assays (Supplementary Figs 1–3). This interaction seemed to be specific because it was abolished by deletion of the first 90 amino acid residues from the Skp2 amino terminus, which contains the D-box motif, or by removal of the Cdh1 WD-40 domain. A Skp2 mutant in which the D box alone was deleted (Skp2 Δ db) showed some binding to Cdh1 in these assays. However, it is known that the D-box motif in another APC substrate, cyclin B, is important for its polyubiquitination by APC, but not for APC binding⁸. Although Skp2 lacks a recognizable KEN-box motif, there are many examples of APC^{CDH1} substrates that contain a D box without a KEN box³.

As a direct test of the role of the D-box motif in APC-mediated degradation, we examined the cell-cycle dependence of Skp2 degradation. We found that Skp2, but not Skp2 Δ db, was rapidly degraded by cell extracts prepared from synchronized G1 cells but not by extracts from cells at the S phase, at which point APC^{CDH1} is inactive (Fig. 1b, c). Moreover, degradation of Skp2 was inhibited in the presence of a recombinant polypeptide corresponding to the N terminus of cyclin B1, which is a very specific competitive APC inhibitor⁹ (Fig. 2a). Earlier studies showed that *Xenopus* interphase cell extracts can support APC^{CDH1}-dependent polyubiquitination and degradation if supplemented with recombinant Cdh1 (ref. 10). Skp2, but not Skp2 Δ db, was efficiently degraded by such extracts (Fig. 2b) in a Cdh1-dependent manner. By contrast, cyclin B, but not Skp2, was destroyed by *Xenopus* extracts under conditions where APC^{CDH1} was active (data not shown). Finally, Skp2, but not Skp2 Δ db, was polyubiquitinated *in vitro* by immunoaffinity-purified APC complexes derived from mammalian (Fig. 2c) or *Xenopus* (Fig. 2d) cell extracts. As expected, polyubiquitination was abolished when methyl-ubiquitin was substituted

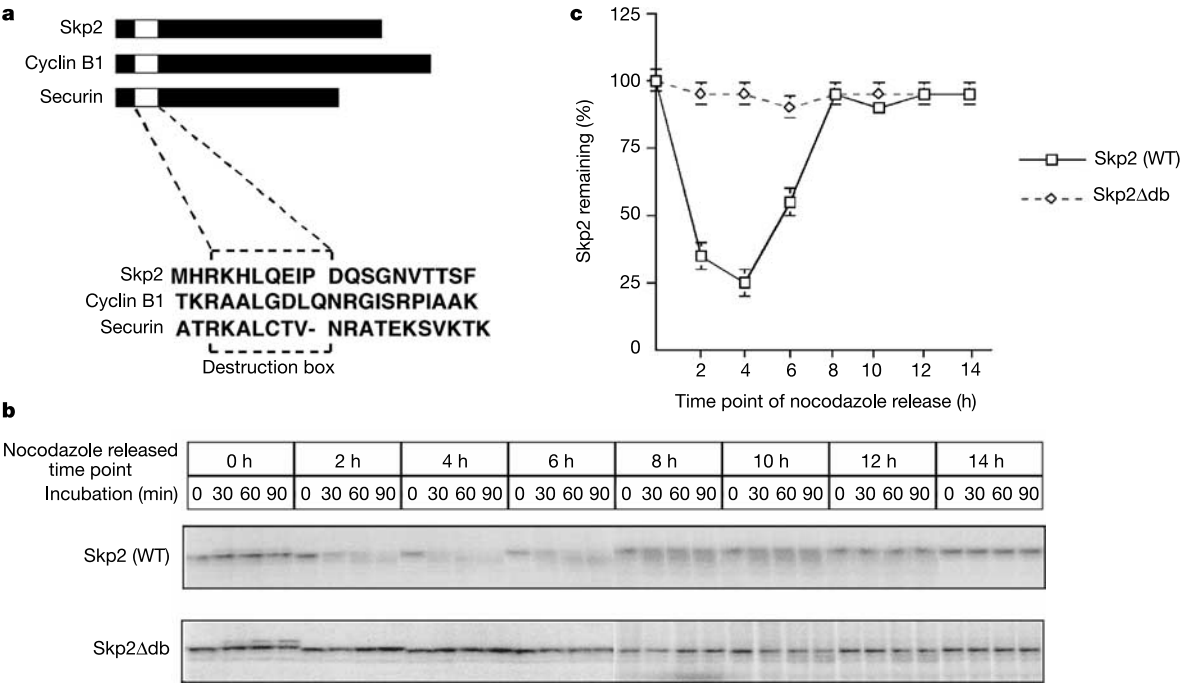


Figure 1 D-box-dependent degradation of Skp2 in G1 extracts. **a**, Alignment of D boxes in Skp2, cyclin B1 and securin. **b**, Autoradiograms of ³⁵S-labelled Skp2 after incubation for 0–90 min in synchronous HeLa cell extracts prepared at the indicated times after nocodazole release. WT, wild type. **c**, Quantification of band intensities in **b**.

for ubiquitin in these reactions (Fig. 2c). Collectively, these data indicate that APC^{CDH1} can polyubiquitinate Skp2 in a D-box-dependent manner.

To investigate whether APC^{CDH1} is a major regulator of Skp2 *in vivo*, we monitored Skp2 levels after experimental reduction of APC activity through depletion of Cdh1. Transfection of HeLa cells with three different short interfering (si)RNAs (out of a total of five siRNAs tested) decreased the steady-state levels of Cdh1 in asynchronously growing HeLa cells (Fig. 3a). Notably, transfection of cells with these three siRNAs increased Skp2 levels compared with the ineffective Cdh1 siRNAs or various control siRNAs. The small fluctuations in Skp2 levels in the absence of Cdh1 downregulation are probably due to siRNA off-target effects.

To exclude the possibility that the observed changes in Skp2 were due to changes in the distribution of the population of cells in the cell cycle, we measured the cell-cycle state by fluorescence-activated cell sorting (FACS) and by immunoblotting for cell-cycle markers. In these experiments, HeLa cells were treated with Cdh1 or control siRNA, synchronized at mitosis with nocodazole, and released into G1. In cells treated with control siRNA, Skp2 levels fell after release and remained depressed until approximately 12 h later, at which point cells began to enter S phase (Fig. 3b, c). p27 levels rose during G1 coincident with the fall in Skp2 levels. By contrast, in cells that were depleted of Cdh1 by siRNA, Skp2 remained at high levels; other APC^{CDH1} substrates, such as Plk and Cdc20, also avoided downregulation. In the continued presence of high levels of Skp2, accumulation of p27 was markedly attenuated. The turnover of some other APC substrates, such as cyclin A, geminin and cyclin B, was minimally affected, perhaps because inhibition of Cdh1 allows Cdc20 to persist—APC^{CDC20} can substitute for APC^{CDH1} for these substrates¹¹ (Fig. 3b; see also Supplementary Fig. 4). These behaviours were clearly observed within 6 h after nocodazole release, before any changes in cell-cycle distribution were apparent by FACS (Fig. 3c). Cells treated with Cdh1 siRNA (but not control siRNA) entered S phase approximately 3 h earlier than the untreated cells, in keeping with previous results obtained with a neutralizing Cdh1 antibody¹². These kinetic studies suggest that persistence of high levels of Skp2 produced by downregulation of Cdh1 is a cause, rather than a consequence, of accelerated cell-cycle progression. In

support of this conclusion, accelerated cell-cycle progression after Cdh1 removal was abrogated by Skp2 siRNA (Fig. 3d, e). Therefore Skp2 is epistatic to Cdh1 in this assay.

In a reciprocal set of experiments, we induced Cdh1 expression in asynchronously growing U2OS cells and this led to marked downregulation of Skp2, in keeping with an earlier report¹². As expected, increased levels of Cdh1 also led to decreased levels of other APC substrates (Fig. 3f). Downregulation of Skp2 in this setting was attenuated by proteasome inhibitors (Fig. 3g). Finally, as a test of the specificity of Skp2 for APC, we found that Skp2Δdb (as with a Skp2 truncation mutant that cannot bind to Cdh1) was degraded much more slowly relative to wild-type Skp2 in HeLa cells entering G1 after nocodazole-induced mitotic arrest (Fig. 4a–c; see also Supplementary Fig. 2b). In these experiments, the amount of Skp2 introduced into the cells was kept low so that proteins required for Skp2 degradation would not become limiting. Under these conditions, exogenous Skp2 does not alter cell-cycle kinetics (Fig. 4d). We conclude that the disappearance of Skp2 in G1 is largely due to polyubiquitination by APC^{CDH1} in a D-box-specific manner.

Skp2 is the substrate recognition module of the SCF^{SKP2} ubiquitin ligase complex and is an important regulator of S-phase entry. As with most key regulators its abundance must be tightly controlled. Loss of Skp2 leads to inappropriate accumulation of p27, c-Myc and cyclin E, and it is thought that this contributes to the associated polyploidy and centrosomal duplication^{13–15}. Conversely, forced expression of Skp2 leads to proliferation and transformation *in vitro* and tumour formation *in vivo*^{16–19}. In keeping with these findings, increased Skp2 levels have been linked to poor prognosis of human cancer patients²⁰. Our findings shed new light on the cellular pathways in G1 for regulating Skp2, and on the importance of APC in G1 control. They also imply that under some circumstances APC^{CDH1} might possess tumour suppressor activity through its inhibition of the Skp2 oncoprotein. In support of these ideas, an accompanying paper reports that ectopic expression of Skp2Δdb in Skp2^{+/+} HeLa cells promotes S-phase entry²¹, and we have observed that ectopic expression of Skp2Δdb, but not wild-type Skp2, blocks neuronal differentiation *in vitro* (N.G.A. and M.W.K., unpublished data).

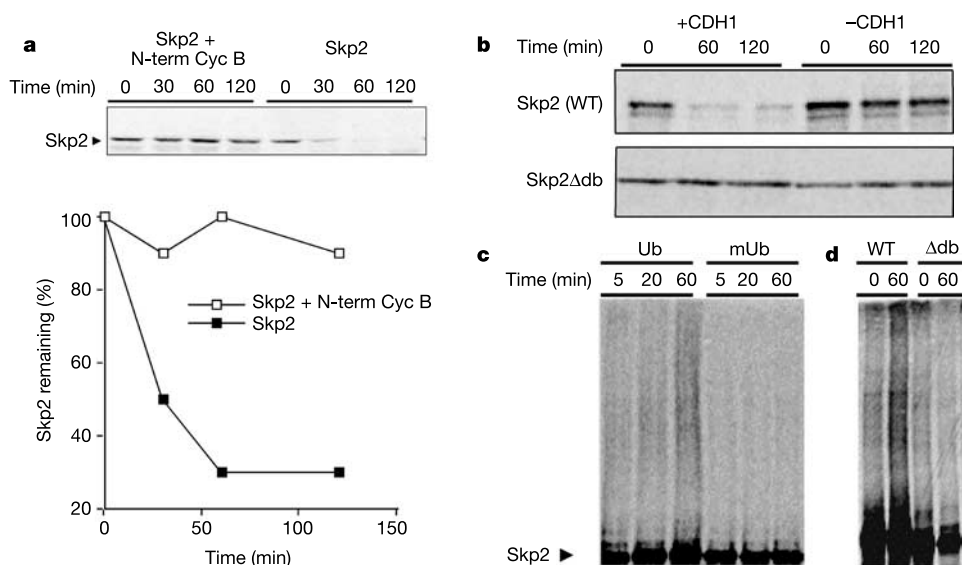


Figure 2 Polyubiquitination and destruction of Skp2 by APC *in vitro*. **a**, **b**, Autoradiograms of ³⁵S-labelled Skp2 after incubation for 0–120 min in HeLa cell extract prepared from G1 cells (**a**) or *Xenopus* interphase extracts that were or were not supplemented with Cdh1 (**b**). An N-terminal fragment of cyclin B1 was added in the reactions in **a** where indicated.

c, **d**, Autoradiogram of ³⁵S-labelled Skp2 after *in vitro* ubiquitination by anti-APC immunoprecipitates derived from HeLa cells (**c**) or *Xenopus* interphase extracts supplemented with Cdh1 (**d**) in the presence of recombinant wild-type ubiquitin (Ub) or methyl-ubiquitin (mUb). All the reactions in **d** contained ubiquitin.

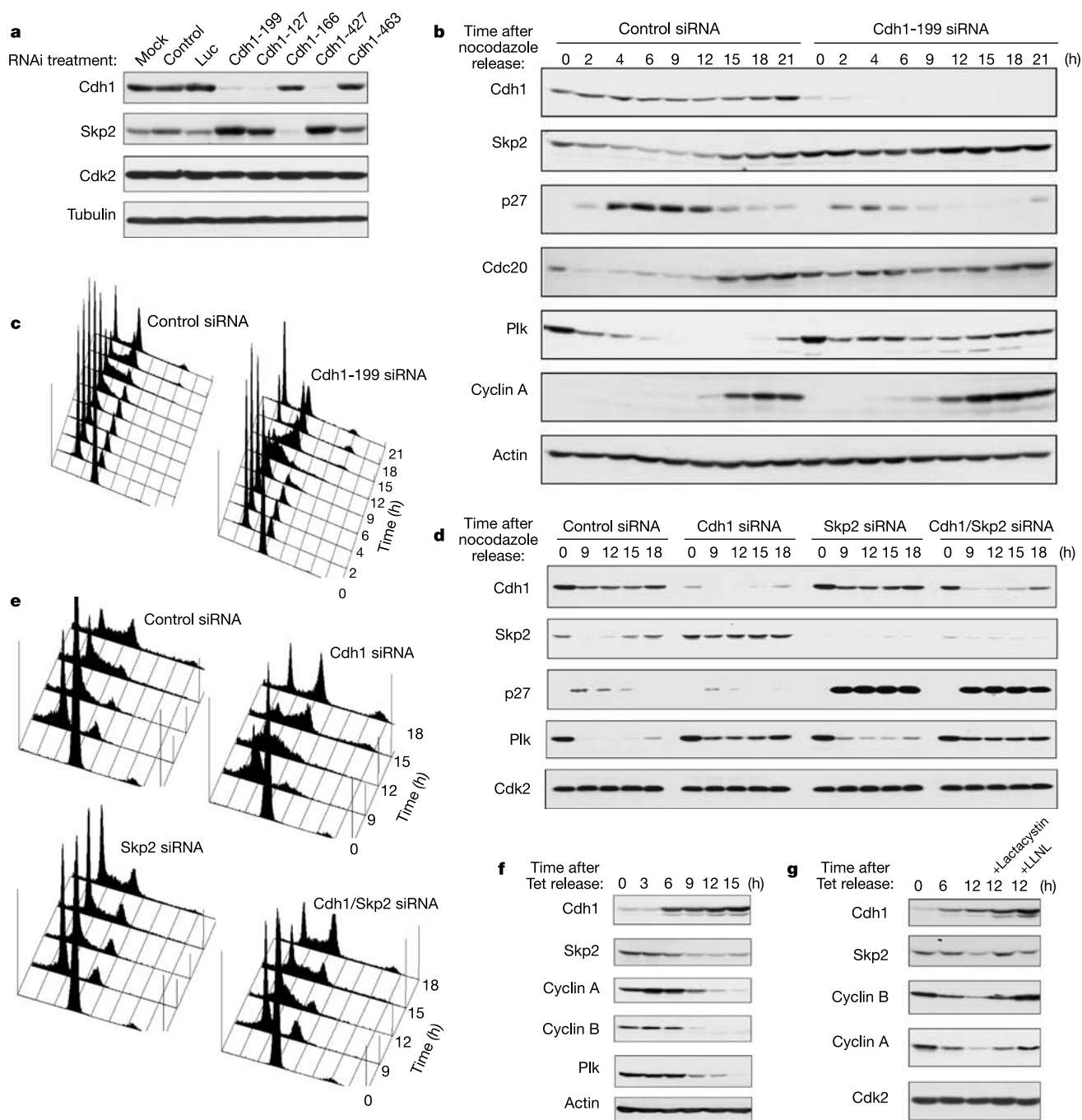


Figure 3 Cdh1 is necessary and sufficient for downregulation of Skp2 in G1.

a, Immunoblot analysis of HeLa cells transfected with the indicated siRNA. The control lane is scrambled cyclin A siRNA; Luc, siRNA directed against firefly Luciferase. siRNA, short interfering RNA. **b–e**, Immunoblot (**b**, **d**) and FACS analysis (**c**, **e**) of HeLa cells transfected with the indicated siRNA, synchronized by growth in nocodazole, and then

released for the indicated periods of time. In **c**, **e**, the y axis indicates cell number and the x axis indicates DNA content as determined by propidium iodide staining. **f**, **g**, Immunoblot analysis of U2OS cells that were engineered to produce Cdh1 after removal of tetracycline (Tet)¹². In **e** the proteasomal inhibitors lactacystin or LLNL were added where indicated.

Although pointing to APC as the key regulator of the stability of Skp2 in G1, our data are not necessarily incompatible with earlier indications that Skp2 may be regulated in quiescent cells by a different ubiquitin protein ligase containing the Skp2-bound SCF components Skp1, Cul1 and Roc1 (also called Rbx1 or Hrt1)⁶. Our study involved cells that had not entered a quiescent period and were proceeding through the cell cycle. Notably, we found that a Skp2 F-box mutant that cannot form SCF complexes is a better

APC^{CDH1} substrate than wild-type Skp2 *in vivo*, possibly owing to competition between APC^{CDH1} and apo-SCF^{SKP2} (data not shown).

Although the cell cycle often seems to be a linear progression of events, once compared to a row of dominoes²², in fact its successful orchestration requires events in one phase to set the tempo or to monitor events in the other phase. The tempo is insured not only by mutually inhibitory phosphorylation events but also by mutually inhibitory degradation events. Thus, in addition to a kinase cycle we

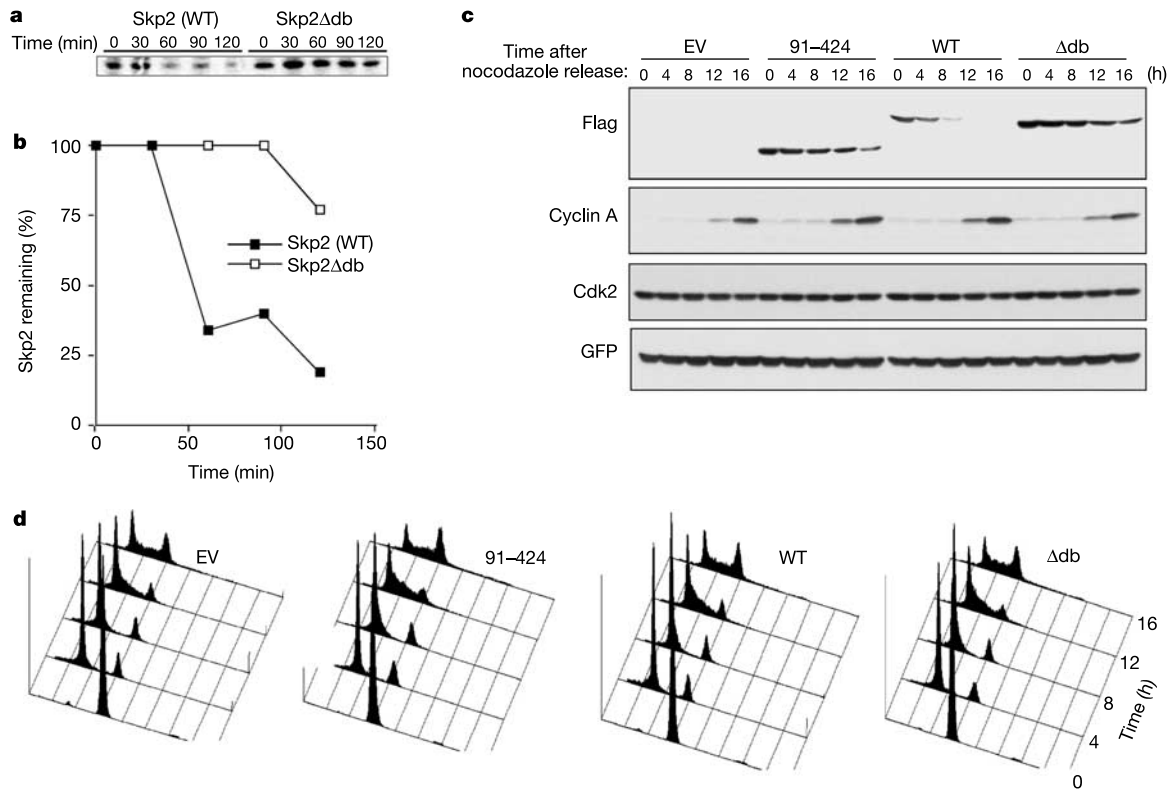


Figure 4 D-box-dependent turnover of Skp2 in G1. **a**, **b**, Pulse-chase analysis of HeLa cells transfected to produce HA-Skp2 (wild type or Δ db). Cells were synchronized in M phase with a thymidine block followed by growth in nocodazole. Nocodazole was absent during the chase. HA-Skp2 was recovered by anti-HA immunoprecipitation and was detected by autoradiography (**a**). **b**, Quantification of band intensities in **a**. **c**, **d**, Anti-Flag

immunoblot (**c**) and FACS analysis (**d**) of HeLa cells transfected to produce limiting amounts of Flag-Skp2 (wild type, Δ db or 91-424), along with a green fluorescent protein (GFP) marker plasmid, synchronized in M phase with nocodazole, and then released for the indicated time periods. See Fig. 3c, e legend for definition of axes in **d**. Transfection efficiency was >50% by GFP. EV, empty vector.

have a degradation cycle, with APC active in M and early G1, and SCF active in late G1, S and early G2 (ref. 1). As APC levels drop in the G1 phase, kinase levels rise; these set in motion the next phase of the cell cycle. The potential importance of events in the M phase that reach into G1 is suggested by our finding that APC prevents accumulation of the SCF component Skp2 in G1 and this allows levels of p27 to rise. Yet another intercycle interaction reaching from G1 into the next mitotic cycle involves another F-box protein, TOME-1, which is also degraded by APC^{CDH1} (ref. 23). Degradation of TOME-1 allows the accumulation of Wee1, which has a role in mitotic entry during the next cycle. As proteolysis via both SCF and APC seems to be central to progression through G1, it will be important to understand the key features of this proteolysis engine, such as when APC activity is terminated in G1 and whether active SCF affects the accumulation of active APC.

Methods

Plasmids

Myc-tagged tagged Cdh1 plasmids²⁴ were a gift of J. Lukas. Skp2 N-terminal and carboxy-terminal deletion mutants were made by polymerase chain reaction (PCR) amplification of the desired regions of a human Skp2 complementary DNA with primers that introduced 5' *Hind*III sites and 3' *Sal*I sites followed by subcloning into pCMV2-Flag cut with these two enzymes. Haemagglutinin (HA)-tagged Skp2 cDNAs were also cloned as *Clal*-*Asc*I fragments into pGEX4T (Pharmacia) and as *Eco*RI-*Xho*I fragments into pCS2.

Antibodies

Anti-Cdh1 antibody (CC43) was from Oncogene. Anti-Skp2 antibody (SC-7164), anti-Skp1 antibody (SC-7163), anti-cyclin A antibody (SC-751), anti-cyclin B antibody (SC-245), anti-Cdc20 antibody (SC-8358), anti-geminin antibody (SC-13015), anti-Plk1 antibody (SC-17783), anti-Cdk2 antibody (SC-163), anti-p27 antibody (SC-528), anti-Myc antibody (SC-40), anti-Cdc27 antibody (AF3.1), polyclonal anti-actin antibody (SC-1615), polyclonal anti-HA antibody (SC-805) and peroxidase-conjugated anti-goat

secondary antibody (SC-2056) were from Santa Cruz. Anti-tubulin antibody (T-5168), polyclonal anti-Flag antibody (F7425), monoclonal anti-Flag antibody (F-3165), peroxidase-conjugated anti-mouse secondary antibody (A4416) and peroxidase-conjugated anti-rabbit secondary antibody (A4914) were from Sigma. Monoclonal anti-HA antibody (MMS-101P) was from Covance.

Cell synchronization

For Figs 3b, d and 4b, adherent HeLa cells were arrested in M phase by growth in 330 nM nocodazole for 18 h, washed, and re-seeded into fresh media. For Fig. 4a, adherent HeLa cells were grown in the presence of thymidine (2 mM) for 18 h and then released into thymidine-free media for 3 h. The cells were then grown in 330 nM nocodazole for 12 h before release into nocodazole-free media.

Skp2 binding assays

Cells were lysed in EBC buffer (50 mM Tris pH 8.0, 120 mM NaCl, 0.5% Nonidet-P40) 48 h after transfection. One milligram of clarified cell extract was incubated with 1–2 μ g of the indicated antibody overnight at 4 °C in a final volume of 750 μ l. Monoclonal anti-actin or rabbit preimmune IgG were used as negative controls. Immune complexes were recovered on protein A Sepharose, washed five times with NETN buffer (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% Nonidet-P40), and eluted by boiling in SDS-containing sample buffer. Bound proteins were resolved by SDS-polyacrylamide gel electrophoresis (PAGE), wet-transferred to a PVDF membrane and immunoblotted with indicated antibodies. Binding to immobilized GST fusion proteins was as described previously²⁵.

Skp2 *in vitro* degradation assays

In vitro degradation assays were as described previously^{26,27}. Briefly, ³⁵S-labelled Skp2 was made by *in vitro* translation using a rabbit reticulocyte lysate (TNT, Promega) and incubated with either HeLa cell extract²⁶ or *Xenopus* interphase egg extract²⁷ at room temperature supplemented with ubiquitin (0.1 μ g ml⁻¹), cycloheximide (1 μ g ml⁻¹) and an energy-regenerating system²⁸. Where indicated, recombinant-cyclin B residues 1–100 (final concentration 100 μ M)⁹ produced in *Escherichia coli* or recombinant Cdh1 (ref. 10) (final concentration 0.4 nM) produced in SF9 cells were added. Reactions were stopped by adding SDS sample buffer.

Skp2 *in vitro* ubiquitination assays

Xenopus interphase extract²⁸ (1 ml) supplemented with recombinant Cdh1 (ref. 10) (final

Cdh1 concentration 0.4 nM) or HeLa cell extract (2 ml)²⁶ was incubated with 4 µg anti-Cdc27 antibody for 4 h at 4 °C. Immune complexes were captured on protein A Sepharose (Biorad) (50 µl), washed four times in XB buffer (20 mM HEPES pH 7.8, 100 mM KCl, 5 mM MgCl₂) supplemented with 1% Triton-X100, washed once with XB buffer with 300 mM NaCl, and once with XB. A total of 5 µl of washed beads were incubated with 1 µl ³⁵S-labelled Skp2, UbCH-10 (2.8 µM) (Boston Biochem), E1 (1 µM) (Boston Biochem) and an ATP-regenerating system²⁸ in a final volume of 7.5 µl at room temperature with shaking. Ubiquitin (Boston Biochem) or methyl-ubiquitin (Boston Biochem) (10 mg ml⁻¹) was also included, as indicated. Aliquots of 1 µl of the ubiquitination reactions were removed, boiled in SDS-containing sample buffer and resolved by SDS-PAGE.

siRNA

siRNAs were purchased from Dharmacon and transfected into subconfluent cells using Oligofectamine (Invitrogen) according to the manufacturer's instructions. Cdh1 siRNAs were named according to the first Cdh1 codon in their recognition sites and corresponded to Cdh1 nucleotide sequences 266–286 (Cdh1-127), 305–325 (Cdh1-166), 338–358 (Cdh1-199), 566–586 (Cdh1-427) and 602–622 (Cdh1-463). The Cdh1-199 siRNA²⁹ and Skp2 siRNA³⁰ were described previously. An siRNA corresponding to a 'scrambled' cyclin A sequence (sense strand 5'-AACAAACUGCCCGACGGAA-3') and a firefly luciferase siRNA served as negative controls.

Pulse-chase analysis

After transfection and synchronization, cells were grown in cysteine/methionine-free media (Invitrogen) for 20 min and labelled in media supplemented with trans-label (50 µl per 100-mm plate, 14 mCi ml⁻¹, ICN) for an additional 20 min. Cells were washed twice with PBS and chased with DMEM supplemented with 10% fetal bovine serum, 2 mM cysteine and 2 mM methionine. At various time points thereafter cell extracts were prepared in PBS containing 1% Triton-X100 and incubated with anti-HA antibody coupled to protein A sepharose (Santa Cruz) for 24 h at 4 °C. Immunoprecipitates were washed four times in PBS 1% Triton-X100, once in RIPA buffer, and analysed by SDS-PAGE.

Received 24 November 2003; accepted 27 January 2004; doi:10.1038/nature02381.

- Nakayama, K. I., Hatakeyama, S. & Nakayama, K. Regulation of the cell cycle at the G1–S transition by proteolysis of cyclin E and p27Kip1. *Biochem. Biophys. Res. Commun.* **282**, 853–860 (2001).
- Peters, J. M. The anaphase-promoting complex. Proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
- Harper, J. W., Burton, J. L. & Solomon, M. J. The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev.* **16**, 2179–2206 (2002).
- Zhang, H., Kobayashi, R., Galaktionov, K. & Beach, D. p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase. *Cell* **82**, 915–925 (1995).
- Imaki, H. *et al.* Cell cycle-dependent regulation of the Skp2 promoter by GA-binding protein. *Cancer Res.* **63**, 4607–4613 (2003).
- Wirbelauer, C. *et al.* The F-box protein Skp2 is a ubiquitylation target of a Cul1-based core ubiquitin ligase complex: evidence for a role of Cul1 in the suppression of Skp2 expression in quiescent fibroblasts. *EMBO J.* **19**, 5362–5375 (2000).
- Zachariae, W. & Nasmyth, K. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes Dev.* **13**, 2039–2058 (1999).
- Schwab, M., Neutzner, M., Mockler, D. & Seufert, W. Yeast Hct1 recognizes the mitotic cyclin Clb2 and other substrates of the ubiquitin ligase APC. *EMBO J.* **20**, 5165–5175 (2001).
- McGarry, T. J. & Kirschner, M. W. Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043–1053 (1998).
- Fang, G., Yu, H. & Kirschner, M. W. Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Mol. Cell* **2**, 163–171 (1998).
- Pfleger, C. M., Lee, E. & Kirschner, M. W. Substrate recognition by the Cdc20 and Cdh1 components of the anaphase-promoting complex. *Genes Dev.* **15**, 2396–2407 (2001).
- Sorensen, C. S. *et al.* Nonperiodic activity of the human anaphase-promoting complex-Cdh1 ubiquitin ligase results in continuous DNA synthesis uncoupled from mitosis. *Mol. Cell Biol.* **20**, 7613–7623 (2000).
- Nakayama, K. *et al.* Targeted disruption of Skp2 results in accumulation of cyclin E and p27(Kip1), polyploidy and centrosome overduplication. *EMBO J.* **19**, 2069–2081 (2000).
- von der Lehr, N. *et al.* The F-box protein Skp2 participates in c-Myc proteasomal degradation and acts as a cofactor for c-Myc-regulated transcription. *Mol. Cell* **11**, 1189–1200 (2003).
- Kim, S. Y., Herbst, A., Tworowski, K. A., Salghetti, S. E. & Tansey, W. P. Skp2 regulates Myc protein stability and activity. *Mol. Cell* **11**, 1177–1188 (2003).
- Sutterluty, H. *et al.* p45SKP2 promotes p27Kip1 degradation and induces S phase in quiescent cells. *Nature Cell Biol.* **1**, 207–214 (1999).
- Latres, E. *et al.* Role of the F-box protein Skp2 in lymphomagenesis. *Proc. Natl Acad. Sci. USA* **98**, 2515–2520 (2001).
- Nelsen, C. J. *et al.* Induction of hepatocyte proliferation and liver hyperplasia by the targeted expression of cyclin E and skp2. *Oncogene* **20**, 1825–1831 (2001).
- Gstaiger, M. *et al.* Skp2 is oncogenic and overexpressed in human cancers. *Proc. Natl Acad. Sci. USA* **98**, 5043–5048 (2001).
- Bloom, J. & Pagano, M. Deregulated degradation of the cdk inhibitor p27 and malignant transformation. *Semin. Cancer Biol.* **13**, 41–47 (2003).
- Bashir, T., Dorrello, N., Amador, V., Guardavaccaro, D. & Pagano, M. Control of the SCF^{Skp2-Cks1} ubiquitin ligase by the APC/C^{Cdh1} ubiquitin ligase. *Nature* **428**, 190–193 (2004).
- Murray, A. W. & Kirschner, M. W. Dominoes and clocks: the union of two views of the cell cycle. *Science* **246**, 614–621 (1989).
- Ayad, N. G. *et al.* Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. *Cell* **113**, 101–113 (2003).
- Sorensen, C. S. *et al.* A conserved cyclin-binding domain determines functional interplay between anaphase-promoting complex-Cdh1 and cyclin A-Cdk2 during cell cycle progression. *Mol. Cell Biol.* **21**, 3692–3703 (2001).

- Kaelin, W. G., Pallas, D. C., DeCaprio, J. A., Kaye, F. J. & Livingston, D. M. Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product. *Cell* **64**, 521–532 (1991).
- Wan, Y., Liu, X. & Kirschner, M. W. The anaphase-promoting complex mediates TGF-β signaling by targeting SnoN for destruction. *Mol. Cell* **8**, 1027–1039 (2001).
- Pfleger, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* **14**, 655–665 (2000).
- Murray, A. Cell cycle extracts. *Methods Cell Biol.* **36**, 581–605 (1991).
- Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
- Mendez, J. *et al.* Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol. Cell* **9**, 481–491 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank W. Harper, J. Lukas and M. Meyerson for critical reading of the manuscript; J. Lukas for reagents; M. Pagano for sharing unpublished data; and members of the Kirschner and Kaelin Laboratories for useful discussions. W.G.K. is a Howard Hughes Medical Institute Investigator. This work is supported in part by NIH grants to M.W.K. and W.G.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W.K. (marc@hms.harvard.edu) or W.G.K. (william_kaelin@dfci.harvard.edu).

Insight into tubulin regulation from a complex with colchicine and a stathmin-like domain

Raimond B.G. Ravelli^{1*}, Benoît Gigant^{2*}, Patrick A. Curmi³, Isabelle Jourdain³, Sylvie Lachkar³, André Sobel³ & Marcel Knossow²

¹European Molecular Biology Laboratory (EMBL), Grenoble Outstation, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France

²Laboratoire d'Enzymologie et Biochimie Structurales, UPR 9063, Centre National de la Recherche Scientifique, Bâtiment 34, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

³U440 INSERM/UPMC, Institut du Fer à Moulin, 17 rue du Fer à Moulin, 75005 Paris, France

* These authors contributed equally to this work

Microtubules are cytoskeletal polymers of tubulin involved in many cellular functions. Their dynamic instability is controlled by numerous compounds and proteins, including colchicine¹ and stathmin family proteins^{2,3}. The way in which microtubule instability is regulated at the molecular level has remained elusive, mainly because of the lack of appropriate structural data. Here, we present the structure, at 3.5 Å resolution, of tubulin in complex with colchicine and with the stathmin-like domain (SLD) of RB3. It shows the interaction of RB3-SLD with two tubulin heterodimers in a curved complex capped by the SLD amino-terminal domain, which prevents the incorporation of the complexed tubulin into microtubules. A comparison with the structure of tubulin in protofilaments⁴ shows changes in the subunits of tubulin as it switches from its straight conformation to a curved one. These changes correlate with the loss of lateral contacts and provide a rationale for the rapid microtubule depolymerization characteristic of dynamic instability. Moreover, the tubulin–colchicine complex sheds light on the mechanism of colchicine's activity: we show that colchicine binds at a location where it prevents curved tubulin from adopting a straight structure, which inhibits assembly.

Tubulin, an αβ heterodimer initially identified as the cellular colchicine-binding protein^{5,6}, is the target of numerous small-molecule ligands that interfere with microtubule dynamics, several

of which are of clinical use, in particular for cancer treatment. Most of them bind to one of three different sites: the colchicine, vinblastine and taxol sites. Whereas the taxol site has been identified in the structure of tubulin protofilaments in Zn-sheets⁴, the locations of the other two sites are still uncertain (see, for example, ref. 7). Tubulin also interacts with regulatory cellular proteins, among which stathmin^{2,8}, or each SLD of its related vertebrate proteins, sequesters two tubulin heterodimers^{3,9,10}.

We previously reported a partial structure of the tubulin:RB3-SLD complex (T2R)¹¹, determined on the basis of the structure of the tubulin heterodimer in Zn-sheets. This defined the complex as a head-to-tail assembly of two tubulin molecules whose curvature is maintained by a long RB3-SLD α -helix modelled as a ~ 90 -residue poly-Ala α -helix whose direction could not be assigned. The current tubulin–colchicine:RB3-SLD complex structure has been determined at higher resolution using experimental phases to avoid model bias¹². It allows us (1) to trace and refine the RB3-SLD polypeptide chain in interaction with tubulin along nearly its entire length, (2) to rebuild tubulin heterodimers and compare their structures in a curved, soluble assembly and in protofilaments and (3) to define the interactions of tubulin with colchicine.

In the complex, RB3-SLD adopts a hook-like shape accommodating the two bound tubulin heterodimers (Fig. 1a). RB3-SLD comprises three domains: an N-terminal ‘cap’ domain (residues 4 to 28) that is conserved in the stathmin family, a variable linker domain (residues 29 to 45) and a conserved, mostly α -helical

carboxy-terminal domain (residues 46 to 145)¹³. Within the N-terminal cap domain, residues 7 to 23 form a β -hairpin (Fig. 1b) that extends the β -sheet of the intermediate domain⁴ of the α -tubulin subunit located at one end of the T2R complex. This extension, together with nearby interactions (Fig. 1c), contributes significantly to the stability of the tubulin–colchicine:RB3-SLD complex. Interestingly, Ser 16, which is conserved in the stathmin family and phosphorylated in stathmin in response to a number of signals¹⁴, is located in the tight turn of the β -hairpin (Fig. 1c) and its phosphorylation could inhibit proper formation of the N-terminal cap structure. Furthermore, the residues of tubulin interacting with the N-terminal cap mediate longitudinal contacts between tubulin heterodimers assembled in protofilaments¹⁵. The SLD would therefore sterically hinder the incorporation of the T2R complex at the (+)-end of microtubules via its N-terminal cap (Fig. 1b), in addition to maintaining the curvature via its C-terminal helix¹¹. The capping interaction also explains why RB3-SLD (data not shown) and stathmin, but not an N-terminally truncated form, prevent the formation of tubulin rings^{9,16}. The SLD linker domain involves a proline-rich region that diverges most among SLDs in the stathmin family^{3,13}. This region has no regular secondary structure and is the least ordered part of RB3-SLD in the complex.

Finally, in the RB3-SLD C-terminal domain, electron density for the largest side chains together with three selenomethionine markers permitted the unambiguous assignment of the helical register. In the structure, the majority of the side-chains that

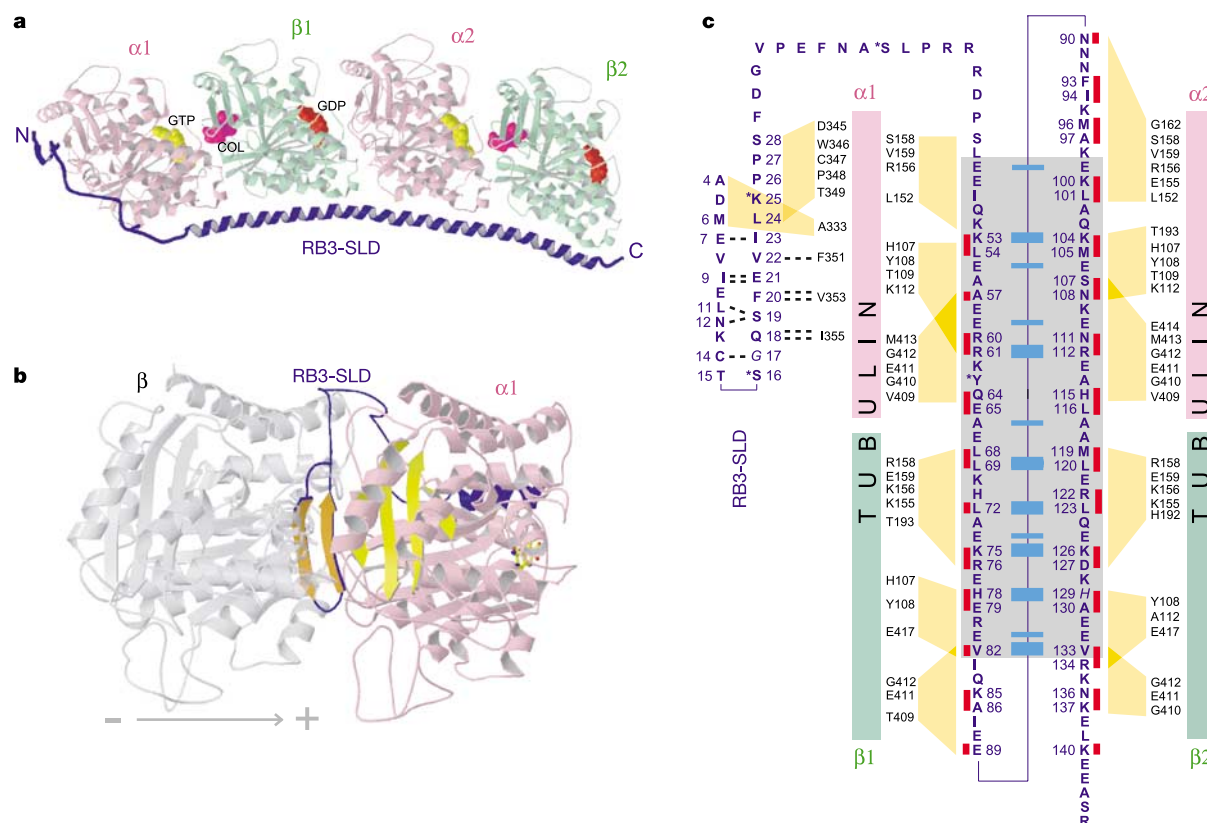


Figure 1 The tubulin–colchicine:RB3-SLD complex. **a**, The complex includes two tubulin $\alpha\beta$ heterodimers, with colchicine bound to β subunits at the interface with α . The RB3-SLD connecting region (residues 29–45) is from tubulin–podophyllotoxin:RB3-SLD, where it is clearest (podophyllotoxin is a competitive inhibitor of colchicine binding to tubulin²³). **b**, The β hairpin (orange) in the N-terminal domain of RB3-SLD caps the T2R complex, extending the β sheet (yellow) of the intermediate domain in the $\alpha 1$ subunit. The extensive overlap with a protofilament (+)-end β subunit¹⁵, preventing the addition of the

T2R complex to a microtubule, is illustrated. **c**, Interactions of RB3-SLD residues with tubulin (except for the least-well-defined RB3-SLD connecting region and for the extension of the tubulin intermediate domain β -sheet) represented by yellow connecting areas. Red bars, residues of the α -helix pointing towards tubulin. Dashed lines, main-chain hydrogen bonds in the extension of the intermediate domain β -sheet. Within the internal repeat (grey), identical residues are connected in blue (thick blue, side-chain pointing towards tubulin). Asterisks indicate positions of stathmin phosphorylation sites.

point towards tubulin are defined and most of the previously proposed tubulin residues that interact with stathmin are confirmed¹¹ (Fig. 1c). The interacting SLD residues include the residues of the mostly hydrophobic seam proposed to be responsible for the interaction¹⁷, but also at least as many polar/charged residues. The data also show that, as hypothesized previously¹¹, each sequence of the repeated pair present in the C-terminal α -helix is identically positioned relative to the corresponding tubulin heterodimer. Remarkably, side chains of six of the 13 repeated residue pairs point away from tubulin (Fig. 1c), raising the possibility of a function of the duplicated sequence and these conserved residues in addition to tubulin binding.

A comparison of the structure of tubulin heterodimers in T2R (which is similar to that of GDP-tubulin in curved protofilaments¹¹) and in straight protofilaments demonstrates three important differences. First, a rotation of $11.6^\circ \pm 1.0^\circ$ is required to superimpose their α and β subunits in the T2R complex. This is not affected by the presence of colchicine, because this rotation was unchanged in lower-resolution diffracting crystals that did not include colchicine.

Second, the relative orientation of the three tubulin subunit domains¹⁸—the N-terminal nucleotide-binding domain, the intermediate domain and the C-terminal domain made of two antiparallel α -helices—is different. The nucleotide-binding and C-terminal domains of each subunit form a rigid block: the root-mean-square deviation (r.m.s.d.) of C α positions in secondary structure elements of these domains in the curved (this work) and straight⁴ structures is 0.9 Å. In both α and β , however, an 8° and 11° rotation, respectively, of their intermediate domains is required to superimpose them in the two structures (the r.m.s.d. of the positions of C α atoms in secondary structure elements after superposition is 0.8 Å) (Fig. 2a).

Third, there are noticeable local differences between the curved and straight structures. They comprise conformational changes in loops located at longitudinal interfaces in protofilaments¹⁵ and also changes in the long, flexible, M and H1–S2 loops, mostly due to the influence of crystal packing (for the nomenclature of tubulin secondary structure elements and loops, see ref. 4). Changes at the intradimer α – β interface comprise a movement of the β subunit T7 loop and H8 helix, which lie close to the colchicine-binding site (the r.m.s.d. of C α positions after superimposition of the β subunit intermediate domains is 4.9 Å, maximum deviation is 6.6 Å), as well as differences in the conformations of the α subunit T5 and H6–H7 loops. The latter accompanies a 1.5 Å translation of the H7 helix along its axis. Changes between the curved and straight structures at the interdimer α – β interface are even more extensive, with two major differences, both in the β subunit: a different T5 loop conformation and a movement in one block of the H6–H7 loop following a 2.5 Å translation of the H7 helix along its axis, dragging along the C-terminal end of helix H6. In the conformation they adopt in the curved T2R structure, the T5 and H6–H7 loops of a subunit would interfere with the neighbouring subunit positioned as across a longitudinal straight protofilament interface (Fig. 2b; Supplementary Fig. 1).

The T5 and H6–H7 loops contribute to the establishment of a kinetic barrier between tubulin straight and curved structures, because they switch between two conformations when this transition happens. The unfavourable interactions of the H6–H7 loop that prevent the straight assembly of tubulin heterodimers will be observed unless the H7 helix switches back to its location in the straight structure. This helix connects the nucleotide-binding and intermediate domains; in both the α and β subunits and in the curved and straight structures, residues in its C-terminal half contact β -strands S7, S8 and S10 of the intermediate domain. As a consequence, for the H7 helix to occupy its location in the straight structure, the intermediate domain will rotate towards the nucleotide-binding domain (Fig. 2c). This rotation allows the

tubulin structure in microtubules to be stabilized through lateral interactions mediated in the intermediate domain by the M loop and, to a smaller extent, by the H9 and H10 helices¹⁹.

The movements described here do not directly or solely depend on GTP hydrolysis because they occur both in GTP-bound

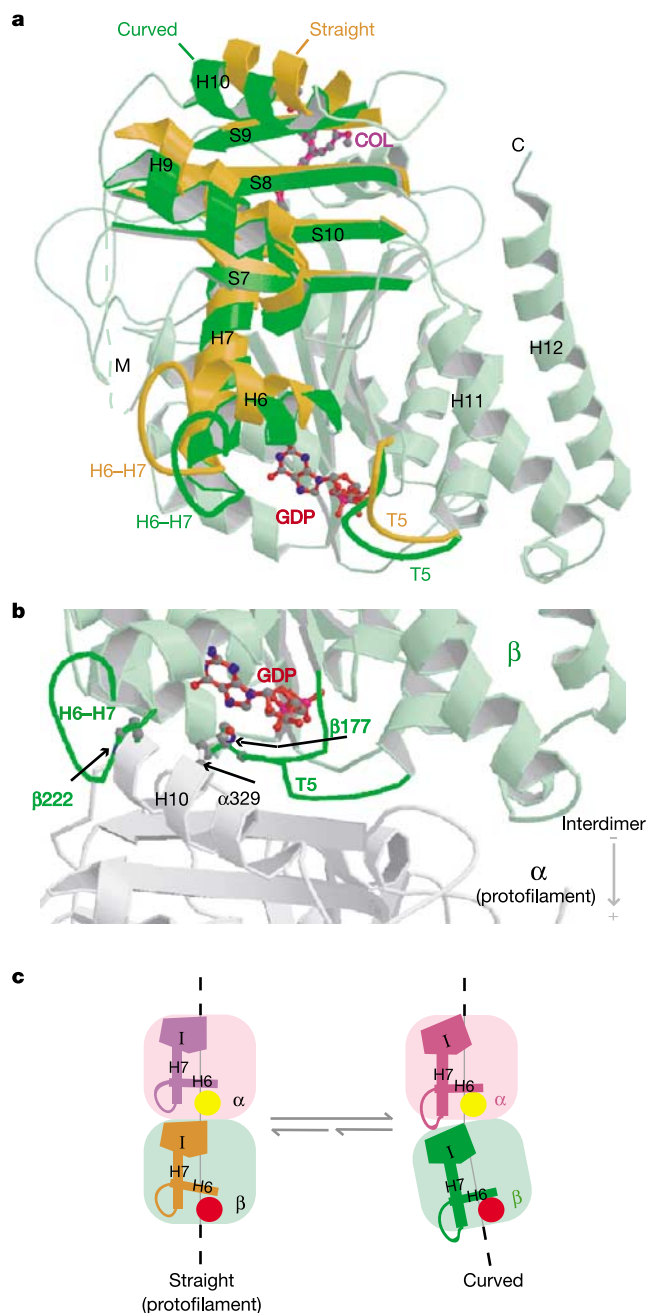


Figure 2 Tubulin subunit changes between the curved and straight structures.

a, The N-terminal nucleotide-binding and C-terminal domains of β subunits in T2R and protofilaments are superimposable. Changes in the intermediate domains secondary structure elements, the H6–H7 helices as well as the T5 and H6–H7 loops (bright colours) in T2R (green) and protofilaments (orange) are shown. Poorly defined M-loop residues that were not traced are presented as a dashed line. **b**, The interference between a T2R β subunit and a protofilament α subunit positioned as across an interdimer longitudinal interface. For clarity only two β subunit residues, β 222 and β 177, which would make strongly disfavoured interactions with helix H10 of the α subunit (with the main-chain and residue α 329, respectively), are presented. **c**, Cartoon representing the displacement of the intermediate domain (I) in α and β during their curved-to-straight conformation switches.

α subunits and GDP-bound β subunits. The coupling of unfavourable longitudinal interactions to the movement of the intermediate domain suggests a reciprocal link between straight tubulin assembly and lateral interactions: lateral interactions force the straight tubulin conformation to be adopted and allow microtubules to assemble and, reciprocally, straight tubulin assembly allows lateral interactions to be established all along microtubule protofilaments.

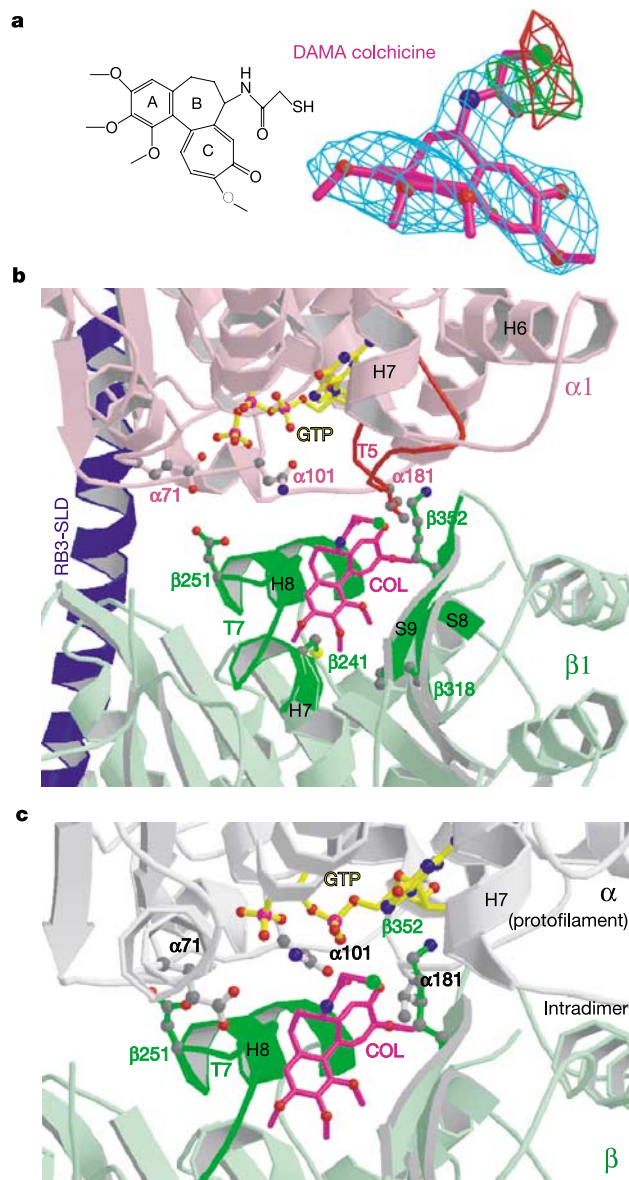


Figure 3 The colchicine-binding site on tubulin. **a**, DAMA-colchicine superimposed on electron-density maps. The $3.5 \text{ \AA}$ $F_{\text{obs}} - F_{\text{calc}}$ omit map (cyan) is contoured at 3σ . The $4 \text{ \AA}$ $F_{\text{obs}} - F_{\text{calc}}$ maps calculated with colchicine (red, contoured at 3σ) and with DAMA-colchicine but F_{obs} of a tubulin-colchicine:RB3-SLD complex (green, contoured at -3.3σ) are also presented. **b**, The colchicine site in the tubulin-colchicine:RB3-SLD complex (bright colours: tubulin loops and secondary structure elements contacting colchicine). **c**, Interference between colchicine binding and the straight conformation of tubulin in protofilaments. An α subunit is positioned near a colchicine-bound β subunit as across an intradimer longitudinal contact in a protofilament. The α -tubulin subunit is prevented from occupying this position because of: (1) steric hindrance between colchicine and residues α 101, α 181 and GTP; and (2) colchicine forcing the T7 loop, H8 helix (for clarity, only side chains of two interfering residues— α 71 and β 251—are presented) and the Lys β 352 side chain to interfere with α -tubulin.

Our observations also suggest that, in addition to a loss of lateral interaction energy, depolymerization of one protofilament induces a destabilization of the neighbouring one whose intermediate domains it contacts, by letting it curve. Indeed, curvature causes additional losses of lateral (see above) and of longitudinal interaction energy, as longitudinal interactions are much less extensive in curved structures than in microtubules (the longitudinal interface areas are $1,960 \text{ \AA}^2$ and $2,200 \text{ \AA}^2$ in T2R compared with $3,000 \text{ \AA}^2$ in straight protofilaments¹⁵). This will lead to easier onset of microtubule disassembly and to enhanced depolymerization speed of GDP-tubulin, two essential features of dynamic instability.

An unresolved issue is the structural mechanism by which GTP regulates microtubule dynamics. One possibility is that binding of GTP to the tubulin β subunit results in a conformational change that locks the H7 lever helix²⁰ in its straight tubulin structure position through interactions at its N-terminal end. This would be propagated to the α - β interface and possibly to the α subunit owing to an induced movement of the intermediate domain, 'winding up' the tubulin heterodimer into a straighter²¹ or ready-to-be-straightened conformation. In microtubules, lateral interactions would then double-lock the H7 helix through contacts at its C-terminal end with the intermediate domain, as described above. Following GTP hydrolysis, the first lock would be loosened; the tubulin heterodimer would then be maintained in its straight structure thanks to lateral interactions, but ready to curve and disassemble in an auto-catalytic process as soon as these interactions are lost.

The colchicine-binding site in tubulin (Fig. 3) was identified during refinement by inspection of the σ_a -weighted²² $F_{\text{obs}} - F_{\text{calc}}$ electron-density difference maps (highest peak height, colchicine, is 12σ ; first peak at a different location is 6σ). There is one such site on each tubulin heterodimer. We also determined the $4.2 \text{ \AA}$ structure of a ternary tubulin-podophyllotoxin:RB3-SLD complex and found that podophyllotoxin binds to tubulin at the same site as colchicine (Supplementary Fig. 2), consistent with its ability to compete with colchicine binding²³. A complex where colchicine was replaced by *N*-deacetyl-*N*-(2-mercaptoacetyl)-colchicine (DAMA-colchicine) allowed us to establish the location of the *N*-acetyl group and to define unambiguously the colchicine orientation in its asymmetric electron density (Fig. 3a). The colchicine site is mostly buried in the intermediate domain of the β subunit, boxed in by strands S8 and S9, loop T7 and helices H7 and H8. Colchicine also interacts with loop T5 of the neighbouring α subunit (Fig. 3b), consistent with the observation that colchicine stabilizes the tubulin heterodimer²⁴.

The structure correlates well with biochemical data showing that β 318 tubulin variants have a reduced sensitivity to colchicine²⁵, and that colchicine derivatives substituted at the methoxy positions of ring A can be crosslinked with Cys β 241 (ref. 7). These studies were performed in the absence of any SLD, so the location of colchicine in the tubulin:RB3-SLD complex is expected to be very similar to that in tubulin. It is interesting to note that there is a movement of the T7 loop and of the H8 helix in the tubulin-colchicine complex compared with protofilament tubulin, which makes space available for colchicine binding. This movement is one reason that the observed colchicine site differs from previously proposed sites modelled on the basis of the protofilament structure of tubulin⁷.

The effect of colchicine on microtubules has been studied *in vitro* using the preformed tubulin-colchicine complex, which incorporates at the end of microtubules²⁶. At a low concentration of tubulin-colchicine, microtubule growth is hindered and microtubule dynamics is inhibited, whereas at high concentration microtubules depolymerize¹. What is the structural basis for these effects? Microtubule stability requires both longitudinal and lateral interactions between tubulin molecules; the M loops of straight tubulin subunits are the central elements of the latter interactions¹⁵. When tubulin-colchicine is added to a microtubule, the lateral contacts of the subunit at the newly formed end of the protofilament are not

established because the M loop is displaced (by more than 9 Å in the structure). This is because the straight tubulin conformation is not adopted, as it would lead to a steric clash of colchicine with the α subunit (Fig. 3c). The microtubule mass will stay preserved as long as the proportion of missing lateral contacts remains small. This will be the case at low concentrations of added tubulin–colchicine, as observed¹. By contrast, at high tubulin–colchicine concentrations, as the proportion of missing lateral contacts increases, destabilized microtubule ends will lead to disassembly.

In conclusion, our data establish a relationship between the onset and speed of microtubule disassembly and the changes in tubulin subunits when tubulin switches from its straight to its curved conformation. They further provide a molecular description of the interaction of RB3-SLD and colchicine with tubulin. We expect that our results will be useful also for the development of novel therapeutic colchicinoid analogues. □

Methods

Protein purification and crystallization

See Supplementary Information.

Data collection and structure determination

The crystals often cracked upon soaking and were radiation sensitive, so hundreds had to be screened for isomorphous pieces suitable for data collection under cryogenic conditions. All data were collected on ID14-4 at the European Synchrotron Radiation Facility (ESRF). Despite the relatively weak diffraction power of the crystals, excellent data could be obtained, especially at low resolution (10–6 Å). In contrast to several other structures of large complexes, we have chosen a strategy to minimize the presence of heavy atoms and carefully examined the decay of anomalous signals during data collection. Where possible, data were corrected for radiation damage²⁷.

Experimental phases were obtained using a 3.5 Å native data set, a 4.0 Å SeMet SAD (single-wavelength anomalous dispersion) data set collected at 15 K, a 3.8 Å PIP SAD data set and a 4.5 Å Yb MAD (multiple-wavelength anomalous dispersion) data set. Data set statistics are given in Supplementary Table 1. The program SHARP²⁸ was used for phasing, aided by an iteration of phasing and model refinement, using model phases as external phases in the refinement of the heavy atom sites. The asymmetric unit contains two tubulin heterodimers and one molecule of RB3-SLD. The experimental phases were further improved by density-averaging the two α - and two β -tubulin subunits separately, using the program DM²². The resulting 3.5 Å experimental map is of excellent quality (Supplementary Fig. 3) and reveals many side chains after sharpening by applying a resolution-dependent factor $\exp(-Bd^2)$, where d is the resolution and $B = -50 \text{ Å}^2$ (ref. 22).

Model building and refinement

The register of the RB3-SLD sequence and electron density was unambiguously determined using the SeMet sites as anchors and verified by the presence of density of larger side chains such as Tyr 63, His 78 and His 129. The program REFMAC^{22,29} was used for restrained positional refinement with a maximum-likelihood target function including experimental phase information. About 5% of the residues are currently modelled as alanines. The two α and two β subunits were restrained separately by non-crystallographic symmetry. The R -factor (R_{free}) is 23% (25%) for all reflections between 20 and 3.5 Å and $I/\sigma_1 > -3$. An example of the final map is in Supplementary Fig. 1b. The position of the sulphur atom in DAMA–colchicine was ascertained by refining a model with DAMA–colchicine using diffraction data from a complex containing colchicine and locating a negative peak in the $F_{\text{obs}} - F_{\text{calc}}$ map close to the location of the sulphur atom (Fig. 3a).

Structure analysis

The angle of the rotation that superimposes tubulin subunits in the T2R complex was determined by least squares using coordinates of tubulin:RB3-SLD complexes determined from independently collected diffraction data sets. The value given is the average of 20 determinations, using ten data sets. The largest conformationally invariant fragment in tubulin subunits when comparing the straight and curved structures was identified using a genetic algorithm³⁰.

Received 12 December 2003; accepted 4 February 2004; doi:10.1038/nature02393.

1. Panda, D., Daijo, J. E., Jordan, M. A. & Wilson, L. Kinetic stabilization of microtubule dynamics at steady state *in vitro* by substoichiometric concentrations of tubulin–colchicine complex. *Biochemistry* **34**, 9921–9929 (1995).

2. Belmont, L. D. & Mitchison, T. J. Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. *Cell* **84**, 623–631 (1996).
3. Charbaut, E. *et al.* Stathmin family proteins display specific molecular and tubulin binding properties. *J. Biol. Chem.* **276**, 16146–16154 (2001).
4. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).
5. Weisenberg, R. C., Borisy, G. G. & Taylor, E. W. The colchicine-binding protein of mammalian brain and its relation to microtubules. *Biochemistry* **7**, 4466–4479 (1968).
6. Bane Hastie, S. Interactions of colchicine with tubulin. *Pharmacol. Ther.* **51**, 377–401 (1991).
7. Bai, R. *et al.* Mapping the binding site of colchicinoids on β -tubulin. *J. Biol. Chem.* **275**, 40443–40452 (2000).
8. Cassimeris, L. The oncoprotein 18/stathmin family of microtubule destabilizers. *Curr. Opin. Cell Biol.* **14**, 18–24 (2002).
9. Jourdain, L., Curmi, P., Sobel, A., Pantaloni, D. & Carlier, M.-F. Stathmin is a tubulin sequestering protein which forms a ternary T2S complex with two tubulin molecules. *Biochemistry* **36**, 10817–10821 (1997).
10. Curmi, P. A. *et al.* The stathmin tubulin interaction *in vitro*. *J. Biol. Chem.* **272**, 25029–25036 (1997).
11. Gigant, B. *et al.* The 4 Å X-ray structure of a tubulin:stathmin-like domain complex. *Cell* **102**, 809–816 (2000).
12. Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72–77 (1996).
13. Maucuer, A., Moreau, J., Mechali, M. & Sobel, A. Stathmin gene family: phylogenetic conservation and developmental regulation in *Xenopus*. *J. Biol. Chem.* **268**, 16420–16429 (1993).
14. Beretta, L., Dobransky, T. & Sobel, A. Multiple phosphorylation of stathmin: identification of four sites phosphorylated in intact cells and *in vitro* by cyclic-AMP dependent protein kinase and p34cdc2. *J. Biol. Chem.* **268**, 20076–20084 (1993).
15. Nogales, E., Whittaker, M., Milligan, R. A. & Downing, K. H. High-resolution model of the microtubule. *Cell* **96**, 79–88 (1999).
16. Steinmetz, M. O. *et al.* Op18/stathmin caps a kinked protofilament-like tubulin tetramer. *EMBO J.* **19**, 572–580 (2000).
17. Honnappa, S., Cutting, B., Jahnke, W., Seelig, J. & Steinmetz, M. O. Thermodynamics of the Op18/stathmin–tubulin interaction. *J. Biol. Chem.* **278**, 38926–38934 (2003).
18. Nogales, E., Downing, K. H., Amos, L. A. & Löwe, J. Tubulin and FtsZ form a distinct family of GTPases. *Nature Struct. Biol.* **5**, 451–458 (1998).
19. Li, H., DeRosier, D. J., Nicholson, W. V., Nogales, E. & Downing, K. H. Microtubule structure at 8 Å resolution. *Structure* **10**, 1317–1328 (2002).
20. Löwe, J. & Amos, L. A. Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203–206 (1998).
21. Müller-Reichert, T., Chrétien, D., Severin, F. & Hyman, A. A. Structural changes at microtubule ends accompanying GTP hydrolysis: Information from a slowly hydrolyzable analogue of GTP, guanylyl (α,β)methylene diphosphonate. *Proc. Natl Acad. Sci. USA* **95**, 3661–3666 (1998).
22. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Wilson, L., Bamberg, J. R., Mizel, S. B., Grisham, L. M. & Creswell, K. M. Interaction of drugs with microtubule proteins. *Fed. Proc.* **33**, 158–166 (1974).
24. Shearwin, K. E. & Timasheff, S. N. Effect of colchicine analogues on the dissociation of $\alpha\beta$ tubulin into subunits: the locus of colchicine binding. *Biochemistry* **33**, 894–901 (1994).
25. Burns, R. G. Analysis of the colchicine-binding site of β -tubulin. *FEBS Lett.* **297**, 205–208 (1992).
26. Skoufias, D. A. & Wilson, L. Mechanism of inhibition of microtubule polymerization by colchicine: inhibitory potencies of unliganded colchicine and tubulin–colchicine complexes. *Biochemistry* **31**, 738–746 (1992).
27. Diederichs, K., McSweeney, S. & Ravelli, R. B. Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr. D* **59**, 903–909 (1992).
28. De La Fortelle, E. & Bricogne, G. in *Methods in Enzymology* (eds Carter, C. W. & Sweet, R. M.) 472–494 (Academic, New York, 1997).
29. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
30. Schneider, T. R. A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Crystallogr. D* **58**, 195–208 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the ESRF for the provision of synchrotron radiation facilities. We thank E. Charbaut for discussions, C. Petosa for critical reading of the manuscript and L. Lebeau for providing us with DAMA–colchicine. This work was supported by grants from the Association pour la Recherche contre le Cancer, the CNRS and the INSERM.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K. (knossow@lebs.cnrs-gif.fr). Atomic coordinates have been deposited in the Protein Data Bank under the accession codes 1SA0 (tubulin–colchicine:RB3-SLD) and 1SA1 (tubulin–podophyllotoxin:RB3-SLD).

China

Messages to China from the West

This collection of articles is unprecedented: a *Nature* supplement that was written for researchers in China and originally published, at the end of last year, in the Chinese language.

The primary motivation for this supplement is our recognition of several positive trends. China is developing fast as a major economic force in the Asian Pacific and in the world at large. This is partly thanks to the natural entrepreneurship of the Chinese people, but also due to their dedicated pursuit of new technologies. Equally important, China has over many years sustained its efforts to turn itself into a world-class scientific power. No one who visits China can fail to be impressed with the results.

But there is a worrying aspect to these trends. Despite such continuous development, and despite major expenditure on science by the Chinese government, China is not yet fulfilling its scientific potential, for reasons that are discussed in these pages. This results not only in a low scientific profile on the world stage, but also in lost opportunities to make the most of what the rest of the world can offer, both scientifically and technologically. This also leads to missed opportunities for China to improve its economy and the quality of life for its citizens.

This supplement is a small but, we hope, significant attempt to rectify this major anomaly in international research. Authors from diverse fields here provide constructive advice and proposals for Chinese scientific and technological development. It is appropriate that they are senior researchers embedded in Western scientific culture. But it was also essential that they have an intimate familiarity with China's research landscape and its opportunities.

In this collection there are forward-looking contributions covering several disciplines, from biomedical research to agriculture. There is also advice that is relevant to all disciplines, about the characteristics required if a culture and system is to become universally recognized as scientifically world-class. Two common themes in these articles are the need for greater engagement by China with researchers in other countries, and the need to take measures that will encourage the very best of Chinese researchers in the West to return to their homeland.

I believe that the distinguished contributors to this supplement, all of them of Chinese descent, successfully bridge Chinese and Western cultures in a fashion that can stimulate scientists and policy-makers in China.

Nature is grateful to the China Bridges International (CBI) Fellowship Fund and the Rockefeller Foundation in New York whose financial support has allowed this English version of the supplement to be published.

Above all, I must thank the authors themselves, who avidly seized this opportunity to provide their views about a country whose scientific future they demonstrably care about so much.

Philip Campbell
Editor-in-Chief, *Nature*

Supplement Editor: **Philip Campbell**
Publisher: **Peter Collins**
Supplement Publisher: **Antoine Bocquet**

Production Editor: **Maria Hodges**
Art Editor: **Martin Harrison**
Production: **Sue Gray**

204 Cultural reflections
Mu-ming Poo

206 Making an impact
Ray Wu

208 The new Silk Road
Kenneth Chien & Luther Chien

210 An embryonic nation
Xiangzhong Yang

213 A case for conservation
Chung-I Wu, Suhua Shi & Ya-ping Zhang

215 Agriculture of the future
T. C. Tso

218 Making big money from small technology
James C. Hsiao & Kenneth Fong

221 Follow your nose
Alice Shih-hou Huang



Can China become a world leader in science? p. 206.



Advice for financing a nanotechnology company p. 218.



Building links between the West and China p. 208.



Replacing traditional values with scientific curiosity p. 204.

Cultural reflections

China's economy is booming and yet its scientific output isn't. Mu-ming Poo explains why.

To great acclaim, Wang Ying-lai and his team at the Institute of Biochemistry in Shanghai announced, in 1965, that they had synthesized biologically active insulin. Similar research was being conducted in labs in the United States and across Europe, but Wang's discovery was well in advance of his rivals, and ushered in a new era of synthetic proteins. This example, unfortunately, is the exception rather than the rule for Chinese science. China has a long way to go to be recognized as a leading scientific country, but rapid progress is being made.

Modern scientific research did not begin in China until the early twentieth century. The founding of the Academia Sinica and its associated research institutes in 1928 signified its beginning, but its activity was seriously disrupted by intermittent wars and political turmoil. Science has had little chance to take hold.

Now, given the soundness of the Chinese economy, the steady increase in the government's funding for basic and applied research, and the general appreciation of the importance of scientific development, the time has come for China to make its presence felt on the international research stage.

I have helped to build several academic programmes in China during the past two decades, and I now believe that the remaining obstacles to Chinese research institutions achieving excellence are cultural rather than economic.

Authority versus creativity

The confucian tradition of respecting customs and hierarchy has cast a long shadow over modern China. Authoritarian rule and political conformity in the past decades have hampered the creation of an

environment that fosters individual creativity. Deference to authority and to existing paradigms is a major barrier to scientific breakthrough.

Science education in China is extensive and rigorous, and has won universal praise. But it takes more than this to cultivate scientists; students should be inspired to pursue knowledge itself, and a habit of raising questions needs to be fostered. Challenges to existing evidence, hypotheses and concepts, however naive, ought to be encouraged and seriously addressed.

Respect for authorities and the spirit of conformity leave their mark on the style of scientific research as well. Research programmes in China often closely follow existing lines of research in the West, using similar paradigms. This often leads to competition at a disadvantage.

Colleagues in China often complain that their results are not appreciated, whereas similar work performed in Western countries is published in high-profile journals.

Strengthening the uniqueness of their work will increase its visibility, as will improving its presentation. At the Institute of Neuroscience in Shanghai, we give regular scientific-writing classes, using drafts of manuscripts to illustrate how to improve clarity and precision.

These skills are important, but ultimately it is confidence and skill in attacking important problems at the forefront of science that will lead to major discoveries and international recognition.

Critical scientific exchange is rarely seen in China, especially in public. Yet open and frank dialogue is urgently needed to make scientific conferences in China not just friendly gatherings but intellectual events that stimulate ideas. Undue courtesy may be indispensable for maintaining the confucian order in a traditional Chinese family, but it is detrimental to research institutions. One way to overcome this might be for the organizers of scientific meetings to begin with the statement that critical or negative comments are to be taken as constructive inputs.

The attitude towards critique is also relevant to the submission of scientific papers to international journals. Critical comments by referees may at first glance seem unfair or

hostile. Researchers would benefit from a more positive approach: it is often useful to reflect upon the comments and then go back to the laboratory bench, rather than sending the paper immediately to a different journal without much improvement. For example, investigators from the Institute of Neuroscience have made great efforts to improve the quality of the work upon rejection of their papers, and this approach has been rewarded by a marked increase in the number of publications in high-quality journals in the past few years.

Essential tension

A lack of well-trained researchers is another drawback in most Chinese institutions. Since the late 1970s, hundreds of thousands of students and researchers have gone abroad for advanced training. A fraction of these people have now returned to China, and they constitute a major driving force in Chinese science. It would be of interest to compare the productivity of these researchers before and after their return. For those who did not perform as well in China, we need to look at why. Is it because of insufficient research funding or the lack of a stimulating environment?

In major research institutions around the world, there is always an 'essential tension' that drives scientists to put their heart and mind into solving scientific problems. This atmosphere is created by a desire to excel, by challenges from surrounding colleagues and students, by competition with scientific peers or simply by the pressure from the 'publish or perish' culture.

An intellectual environment where 'adversity breeds creativity' is critical for scientific discovery and technological innovation. Chinese students and researchers working overseas have earned universal praise for their intelligence and diligence, but it is the competitive environment in which they work that has shaped them into high achievers. I expect that when such an environment is provided in China, important scientific achievements will emerge. Thus the most urgent task in building research institutions in China is the creation of an intellectual atmosphere that is conducive to creative work. Until a large number of returnees can accomplish inter-

Until a large number of returnees can accomplish work that is internationally recognized, a significant flow of scientific talent back to China is unlikely to occur.



Submission to rules and rituals is the foundation of confucian education, yet deference to authority and existing theories is a major barrier to scientific breakthrough.

simply because of poor research performance — a common practice in major research institutions elsewhere. Traditionally the Chinese government has taken care of the entire life of a scientist, from college graduation to retirement, regardless of their performance. The result has been the absence of pressure and a lack of incentive to excel.

Many research centres are now instituting regular reviews of their scientific staff. However, a successful merit-based review system requires objective evaluation of research performance and achievement. This must include the reviewers remaining anonymous — an unfamiliar concept to the Chinese. In China, this process is hampered by the shortage of qualified reviewers in each scientific area. Extensive use of international colleagues, as practised by all major research institutions around the world, will help to solve this problem.

Distinctive goals

The rapid scientific progress in the West poses a formidable challenge to Chinese research institutes. We need to ask whether there are sufficient resolve and resources to compete with Western institutes on major unsolved scientific problems.

Despite the spectacular success of the genome projects, major breakthroughs have mostly originated in small laboratories pursuing their own research interests. The important remaining scientific problems are generally well recognized, yet effective approaches to them remain elusive. Opportunities abound for distinct scientific explorations.

It will take time to develop distinct approaches to science away from mainstream influences, and it requires patience and persistence on the part of individual scientists and scientific administrators.

After decades studying Chinese science and civilization, Joseph Needham¹, a historian of Chinese science, concluded that it was the Chinese form of 'bureaucratic feudalism' that inhibited the rise of modern science. A major challenge for China in the new century is to overcome the past and begin a fully fledged development of its research institutions. ■

Mu-ming Poo is head of the Division of Neurobiology, Department of Molecular and Cell Biology at University of California, Berkeley, California, USA. He is also director of the Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China.

1. Needham, J. *Sci. & Soc.* 28, 235 (1964).

nationally recognized work, a significant flow of scientific talent back to China is unlikely to occur.

Burdens beyond science

As well as providing a conducive intellectual environment, Chinese institutes have two other important tasks: reform of the administrative structure, and establishment of a merit-based system for staff evaluation and resource allocation.

Complaints about administrative constraints are universal, but there is an added hurdle for Chinese scientists. Scientific research, like many other aspects of Chinese society, is under direct government control. Major funding is usually awarded to organized research projects that involve large numbers of investigators and subjects that are clearly defined by the government.

Scientific administrators at all levels have enormous responsibility — and power. These administrators often control resources and give instructions rather than

provide services. Substantial restructuring of the administrative system, including reducing the number and increasing the efficiency of administrative staff, as well as simplifying budgeting and reporting procedures (while retaining reasonable fiscal accountability), will be important.

I am not aware of any research institution in China that has terminated the contract of a scientist simply because of poor research performance.

The recent restructuring at the Shanghai Institutes for Biological Sciences is an interesting example. Here supporting offices of several institutes are integrated into central units, with increased efficiency and reduced staff numbers. If steps can be taken to prevent researchers from feeling alienated from central-office staff and homogeneous office practices overriding unique institute needs, this approach may prove a good model for groups of institutes in related research areas.

The quality of a research institution depends on merit-based appointment, promotion and resource allocation. I am not aware of any research institution in China that has terminated the contract of a scientist



Making an impact

Compared with researchers in the United States, Chinese scientists publish far fewer highly cited papers. Ray Wu believes that this can change.

Despite a recent improvement, the number of high-quality research papers published by scientists from mainland China has remained low. I believe that this is because there are few productive scientists in China, and the level of financial support for basic research is inadequate. The system for evaluating research proposals and distributing funds is far from ideal, and does not promote innovative research. In addition, the education system in most universities does not encourage students to think critically and creatively. Many improvements are urgently needed.

The Cultural Revolution, which ended in 1976, severely devastated research in China. At that time, there were probably fewer than 50 internationally recognized biologists in China. Between 1978 and 1989, when Deng Xiaoping served as the country's leader, substantial reforms and progress were made in almost every area of scientific and economic development. In spite of this progress — which has continued under the leadership of Jiang Zemin — a huge gap still exists between the scientific productivity in China and that of the United States. This is because the baseline levels were so very low in 1978. China needs to increase its efforts to accelerate progress on all fronts.

In 2000, the number of biologists, government-funded labs and the total number of publications were similar in China and the United States. But, the number of high-impact papers published by scientists from China was less than 4% that of the United

States. This failing is, I believe, directly proportional to the small number of productive biologists.

Few productive biologists

The productivity of a biologist can be measured by the number of original research papers published in internationally refereed high-impact journals. An arbitrary and not overly demanding standard would consider a biologist productive if he or she had published at least eight highly cited research papers in the past ten years. I consider such a paper to be one published in a journal with an impact factor of at least two. In addition, among the eight papers, at least one should be published in a journal with an impact factor of five or above.

Using my arbitrary standard, there are now only 500 productive biologists in China¹. By contrast, in the United States, the number of productive biologists of Chinese descent is over 3,000. And, the total number of productive biologists in the United States, which has only a quarter of the population of China, is over 40,000. (These numbers are my own estimates derived from published statistics.)

China's low-level output is related to the inadequate and short-term nature of its research funding. This encourages scientists to work on projects that are likely to produce quick results. So, in most cases, their research lacks novelty and creativity. Consequently, the vast majority of papers are published in

Chinese journals, which are usually not read by scientists of other countries and thus are seldom cited.

There are two solutions to the low number of productive biologists. One is to educate a larger number of biologists in China to think critically and originally. For this, China would need to improve its higher-education system. Currently, only a small number of universities have adopted teaching methods similar to those in the United States and Britain, where students

are encouraged to think critically and creatively. The majority of the universities and graduate schools in China place more emphasis on memorizing and accepting facts than on thinking innovatively and asking questions. I

hope that the Ministry of Education will encourage all universities to improve their methods for training scientists.

The second solution is to attract a large number of productive Chinese biologists now working abroad back to China. Many Chinese scientists would like to return, but the inadequate level of funding, the poor research environment and the absence of an efficient management system discourage them from doing so.

To be attractive to its expatriate scientists, China needs to invest more in its basic research, establish a fair and transparent system to review research proposals, distribute research funds based on the merits of the proposal and promote reform of the

There is a popular saying in China: "Small grants, big review; medium grants, small review; big grants, no review."

infrastructure and administrative systems of the various universities and research institutes. China's annual budget for basic research in biological sciences is approximately two billion yuan (US\$ 242 million), which is about 0.02% of China's gross domestic product (GDP) in 2002. By contrast, the US government spends about \$30 billion on research, or roughly 0.3% of its GDP. In addition, in the United States, another \$30 billion comes from industries and foundations, whereas China has almost no private funding.

China is now an advanced developing country, and its GDP has been increasing by between 6% and 10% every year for the past ten years. There is really no justification for the government to say that China cannot afford to allocate more money to basic research, because we are not talking about absolute amounts but about a percentage of the GDP. The low level of support for basic research reflects a low priority.

I propose that the Chinese government takes immediate action by substantially increasing its annual budget for basic research. I hope that by 2020, the level of funding will be at least 0.1% of GDP, and that the number of biologists who publish high-quality papers will rise to 5,000 — ten times higher than at present.

Evaluation and selection

The system for evaluating research proposals and distribution of funds in China also needs substantial improvement. There are three large organizations in China that fund the majority of research in the biological sciences: the Chinese Academy of Sciences (CAS), the Ministry of Science and Technology (MOST) and the National Natural Science Foundation of China (NNSFC).

The NNSFC has the strongest system of peer review in China, and it is open to all applicants. It has adopted a review system similar to that of the US National Science Foundation, but because the number of qualified reviewers is very low, it is often difficult to find sufficient reviewers who are familiar enough with the area of research to make a sound judgment. I have also been told by several reliable scientists that the larger grants of the NNSFC (and those of CAS and MOST) are often not peer reviewed. Instead, the funds are allocated by high-level administrators. In other cases, some reviewers

choose to help their friends instead of basing their decision on the merit of the proposals. Therefore, even though the procedures adopted by the NNSFC are good, the process has been spoiled by human factors and influences.

There is a popular saying: "Small grants, big review; medium grants, small review; big grants, no review."

Overseas researchers

One way of reducing such problems is to invite qualified reviewers from abroad to serve on the proposal review boards. This would minimize conflicts of interest and lighten the load of the small number of qualified reviewers in China. In doing so, it would make the review process fairer and more transparent.

I am glad that two years ago the NNSFC started inviting highly qualified reviewers from the United States to participate in this process for biological sciences. The review results are without doubt much better than before. I hope that the NNSFC will expand the new process to evaluate proposals in all research areas.

I propose that the government appoints a National Biomedical Science Advisory Committee (NBSAC) to provide guidance for the operation and detailed planning in the procedure of reviewing proposals and in deciding the funding level. At least half of the committee members should be based abroad and should have extensive experience of conducting basic research. The NBSAC would help to establish a National Health Research Board (NHRB) to manage and review research proposals.

In addition, the NBSAC would appoint a special monitoring committee to oversee the activities of the NHRB, the progress of those who receive research funds, and reforms of the infrastructure of the universities or institutes in which the investigators work. After reviewing the progress of the researchers, those who excel should be rewarded: the NHRB could advise universities or institutes which scientists to promote based on performance rather than seniority. Promotion should not include additional administrative responsibilities, and outstanding scientists should be encour-

aged to remain as researchers rather than become administrators.

I also propose that the Chinese government establishes a new type of research project — major programme projects (MPPs). This would allow China to target funds to important research areas. China should establish 80 to 100 major projects within the next ten years. The proposal for one of these projects would need to be jointly submitted by at least three productive biological scientists and target an important area of cutting-edge research.

One important feature of the MPPs is that they should attract outstanding biological scientists from abroad. The budget would be for ten years and would be large enough for each individual MPP to attract up to ten excellent scientists. A long-term commitment by the government to support this would be essential. The initial task of the NHRB would be to review the proposals, and then monitor annual progress.

A bright future

I am optimistic that substantial improvements will be made soon, because I believe that China's new leaders will pay much more attention to the issues and problems discussed in this article than their predecessors did and that they are keen to solve any problems. I am hopeful that we will soon see a much larger budget for basic research and that a truly fair and transparent system for reviewing research proposals will be established. In addition, I hope that the proposed MPPs and system for managing and monitoring research proposals and productivity will be implemented. If this becomes

a reality, China will be able to train a greater number of productive biologists and attract many scientists of outstanding calibre from abroad. The number of productive biologists can be doubled within five years, and increased tenfold by 2020.

Ray Wu is a Liberty Hyde Bailey Professor in the Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.

1. Wu, R. in *Chinese Americans in Science and Technology* (C. C.H. Chen, editor), pp. 37-84 (Outer Sky Press, New York, 2003).

Acknowledgements: I thank X. C. Gu, P. C. Huang, B. Lu, Y. Rao and T. C. Tso, among others, for information and advice.

The new Silk Road

Kenneth Chien and Luther Chien look to the past to inspire biomedical research of the future.

Hope lies in dreams, in imagination and in the courage of those who dare to make dreams into reality.

Jonas Salk

For over 2,000 years, China has been at the heart of Asian society and culture. From painting to porcelain, calligraphy to cuisine, it has been a dominant force in the Far East. Now, China is poised to assume a major role in science and technology, particularly in human biology and molecular medicine.

The Chinese government's dream is for their country to become a global hub for biomedical research. But, as Jonas Salk noted, the fulfilment of such a dream requires more than knowledge itself; it needs imagination and courage.

This is an ambitious goal, but the timing is right. An alarming increase in the levels of cancer and cardiovascular diseases in China have made biomedical research a top priority, and timely investment from the government is creating an infrastructure that will appeal to expatriate scientists.

Arabic influence

Perhaps the most challenging step for Chinese scientists is to become familiar with new discoveries throughout the world. In the past couple of centuries, China has been largely isolated from the latest developments in Western science. But this has not always been the case.

The blue and white porcelain of the Ming dynasty in the fourteenth and fifteenth centuries is widely acknowledged to be the pinnacle of human achievement in ceramics.



Benefits of cultural exchange: Ming dynasty ceramics used pigments from the Middle East.

Before the Ming period, Chinese ceramics were dull and lifeless. This art was transformed by the discovery that cobalt-blue pigment could produce a brightness and intensity of blue that has never been duplicated.

But this success needed the crucially important Silk Road. The pigment arrived in China from the Middle East by the Silk Road, a trail established by merchants during the period when the Mongols ruled China in the thirteenth century. This was the major route for cultural exchange between the two regions. The importance of this exchange can be seen in several of the motifs, shapes and designs (some with Arabic characters, specifically made for the Muslim market) on the most prized ceramics in the Istanbul and Tehran palaces.

Could such profitable cultural exchange happen again? We think so. In this era of molecular medicine, we see the 'new Silk Road' as the sharing of information between Chinese biomedical institutes and those of the United States and other Western nations. This exchange would open up many opportunities, which, like the old Silk Road, would be of benefit to both sides.

One attraction for researchers is the sheer

number of people with cardiovascular disease in China — it is predicted that it will affect over 100 million people in the coming years. Coupled with the relatively low cost of hospitalization, China is ideal for clinical studies of novel therapeutic and diagnostic approaches.

Proper controls

Another advantage to the West, though clearly a disadvantage to China, is that standard clinical cardiovascular care in China does not include the routine use of expensive devices and drugs because of their high cost. A true placebo group to act as a control in drug trials is difficult to find in the West, because of the routine prescription of many drugs, but in China a 'clean' group could be easily found. Pilot clinical studies, while not sufficient to justify formal US Food and Drug Administration approval, could provide the impetus for further clinical testing in Western nations. In exchange for this information, it would be important to ensure overseas investment in China's health-care system.

Unravelling the mechanistic basis of diseases that arise from multiple, independent

LUTHER CHIEN

genetic and environmental variables will require international and interdisciplinary teams of doctors and scientists.

International training programmes should be designed in China to develop a new generation of Chinese graduates, who will serve as a route for the exchange of technology and ideas between different subject areas. The recent establishment of a series of new institutes in China, such as the National Institute of Biological Sciences in Beijing¹, the new biomedical institute in Guangzhou², and the new Institute of Molecular Medicine China at Beijing University could provide an infrastructure for such training programmes.

Other attractions

The burgeoning Chinese space programme could also form the basis for a new era of biomedical engineering. A clear precedent is evident from the number of spin-offs from the US National Aeronautics and Space Administration (NASA) programme. This interdisciplinary approach would capitalize on the strengths of Chinese institutions in engineering, physics, chemistry and mathematics.

China could be attractive to the West because of its expertise in non-human primate research, an area that is currently hampered by cost and ethical issues in Western nations. The availability of established primate colonies should give China a strong position in the ongoing non-human primate genome project.

Mouse models of human disease have led to the identification of many genes and pathways that may be linked to disease. But the usefulness of mouse models is limited because of the great differences

in neural and cardiovascular physiology between mice and humans. Translating these experimental observations to studies in humans will be difficult without first having supporting evidence from large-animal models, preferably non-human primates.

Exploiting the less-than-1% difference in genomic sequence between such primate species and humans could be a major niche for China. Non-human primate studies are currently only performed in a handful of US centres and are rapidly vanishing in the European community, as witnessed

Exploiting the minor difference in genomic sequence between chimpanzees and humans could be a major niche for China.



Guangzhou's newly constructed research institute could be a beacon of interdisciplinary research.

by the lack of support for a Cambridge University primate-research facility³. Partnerships between China's research centres and biotechnology and chemistry companies should lead to cost-effective primate-based data on drug safety that would be difficult to obtain in the United States or Europe.

The other main area of opportunity is a product of China's isolation. Chinese communities away from the main industrial centres may hold the key to the genetic risk factors and biomarkers that make the population prone to common diseases.

The recent discovery by Yi-Han Chen and colleagues of a gene for a rare form of familial atrial fibrillation, a disease that is associated with a high risk for stroke and cardiovascular morbidity, shows the value of studying genes from clinically well-defined Chinese patient populations⁴. Identifying the causes of other diseases will require state-of-the-art clinical

phenotyping, and the development of regional networks of referring physicians that are coordinated by a single centre with world-class expertise in human genetics.

All these projects require investment, and China could use its growing market for clinical drugs as leverage to attract support from a consortium of large pharmaceutical, biotechnology and asset-management funds. Large drug companies are already beginning to establish sites in China, particularly in Shanghai. This trend is driven not only by cost-effectiveness, but also by China's strength in chemical research.

Western pharmaceutical and biotechnology firms that are allowed to tap into the growing clinical market in China should be required to provide a minimal level of sup-

port for national biomedical training and research initiatives. The establishment of international business standards to translate these molecular-medicine initiatives into reality could lay the foundation for a strong Chinese effort in drug discovery and biotechnology. To accomplish this objective, China will need to train business professionals to have direct experience in multiple aspects of biotechnology, ranging from intellectual property management to financing.

Family history

Much of this could be achieved through building biomedical links between China and the West. During the past century, a number of Chinese families have moved back and forth between China and the United States, thereby maintaining close ties between China and the West. Our ancestor Qian Zeng Qi, a provost of Beiyang University, had the foresight to send all his sons for education abroad. It was Qian's dream that his sons would eventually return to China and bring Western scientific and business precepts to mainstream Chinese society.

Today, there are many highly trained Chinese in virtually every area of science, technology and business in the United States, and the conditions are ideal for their return to China.

Capitalizing on this human Silk Road and moving towards globalization will be key if China is to enter the front ranks of molecular medicine. Encouraging close collaboration between overseas and Chinese biomedical scientists should lead to China rapidly becoming a centre for the growing global network of molecular medicine. Perhaps an ancient Chinese parable aptly sums up this philosophy: "If you do not venture into the tiger's den, how can you capture the tiger's cub?"

Kenneth Chien is the director of the University of California, San Diego (UCSD), Institute of Molecular Medicine and adjunct professor of the Salk Institute, UCSD Institute of Molecular Medicine, La Jolla, California, 92093, USA.

Luther Chien is a departmental fellow (retired) of E. I. DuPont De Nemours, Wilmington, Delaware, 19898 USA.

1. Cyranoski, D. *Nature* **420**, 257 (2002).

2. Cyranoski, D. *Nature* **425**, 333 (2003).

3. *Nature* **427**, 471 (2004).

4. Chen, Y. H. et al. *Science* **299**, 251–254, (2003).

An embryonic nation

Liberal views on human-embryo technology make China ideal to become a world leader in this field. Xiangzhong Yang explores its potential.

As China pours money into regenerative medicine, its scientists are starting to explore the field's technical and ethical frontiers¹. Therapeutic cloning, stem-cell studies and other research areas that use animal or human embryos are controversial and raise religious and ethical questions, as well as concerns over animal welfare.

As a result, many Western governments are wary of such research. These issues have led to unsupportive policies for cloning-related research, and the high cost of clinical trials for any proteins developed using this technology have forced many scientists and commercial companies to abandon promising research and to lose out on potentially profitable products².

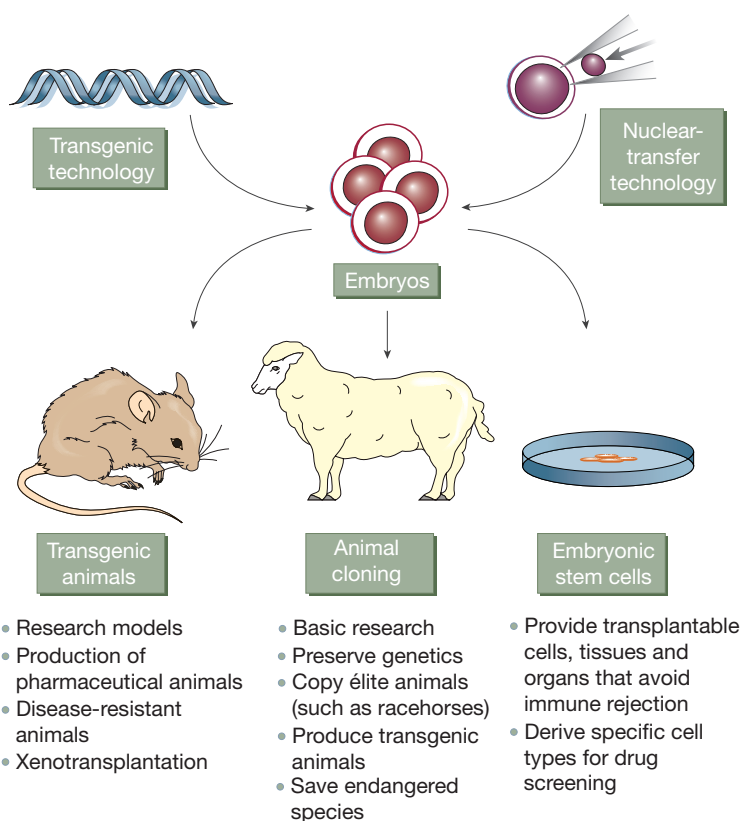
China has a cultural environment with fewer moral objections to the use of embryonic stem cells than many Western countries, and, if it can provide a supportive funding and academic environment, it could take a leading role in this field. These technologies offer unprecedented research and commercialization opportunities for China.

Public approval

The success of cloning using nuclear transfer (the same technique used to produce the first cloned animal — Dolly the sheep)³ shows that the nucleus of a differentiated cell, for example a skin cell, from an adult, can be reprogrammed and converted back to an embryo-like cell that can give rise to all cell types and tissues. This technology is a significant improvement on transgenic-animal production, and offers a fresh approach to agriculture through the cloning of elite bulls or racehorses, as well as to the cloning of endangered species such as pandas and to the creation of animal models of human disease for basic medical research.

China is firmly opposed to reproductive human cloning⁴. But human therapeutic cloning (which is when a patient's own cells are used to form stem cells that grow into tissues needed by the patient) is legally per-

The many uses of embryo technology.



mitted, unlike the current situation in many other countries, particularly the United States and Germany. China has probably the most liberal environment for embryo research in the world: most Chinese media report positively on achievements in transgenics, therapeutic cloning and related embryo-based research. Furthermore, the public shows little opposition to such studies.

This liberal environment and acceptance of these biotechnologies stems from governmental and public recognition of the significance of this work. In addition, the

relatively easy access to human material, including embryonic and fetal tissues, in China is a huge advantage for researchers.

Human therapeutic cloning and animal-based embryo biotechnology research should be a top research priority if China wants to focus on scientific areas with the potential to make her a world leader in a short period of time.

Chinese scientists have already produced transgenic rabbits, goats and cows within a few years of the technology becoming available. Likewise, China has succeeded in cloning goats and cattle and most recently

was part of the first team in the world to successfully clone a rat⁵, which is one of the most difficult species to clone. Other high-profile results have come from Huizhen Sheng's team at Shanghai Second Medical University, which extracted stem cells from embryos by fusing adult human cells with specially prepared rabbit eggs⁶. And Guangxiu Lu's team at Xiangya Medical College in Changsha has cloned human embryos⁷ to multicellular blastocyst stage. This and recent progress in successfully extracting embryonic stem cells⁸ represent significant progress towards human therapeutic cloning.

China has established a good base for human-embryo research, but this field is still in its infancy and far from sufficient to prepare a country to play a leading role in the world. It remains to be seen whether China can build on her strength in the promising area of human therapeutic cloning research to become a world leader.

For instance, China acquired the necessary technologies to produce transgenic animals and pharmaceutical proteins decades ago, but has shown no signs of production, let alone commercialization, of such proteins. Meanwhile, Western companies have entered into various phases of drug development and clinical trials in the past decade, even though in China the drug-approval phase is much shorter (five years rather than roughly 15 years in the United States).

Various factors may contribute to this lack of drug development, but China needs to improve the following: funding for basic research, the efficiency of the technology-transfer system, the return rate of talented individuals from developed countries, and the number and quality of national and international collaborations.

The Chinese government falls far behind more developed countries in commitment to research and development. China only spent US\$12.5 billion (1.1% of its gross domestic product (GDP)) in 2001 (ref. 9). In contrast, the percentage of the GDP allocated in Japan, the United States and South Korea for research and development is 3.0, 2.8 and 2.7, respectively¹⁰ (see the table on the right). In 2003, the US National Institutes of Health alone has a budget of US\$27 billion¹¹, more than dou-



Huizhen Sheng is internationally respected for her stem-cell research.

ble the total that China now spends on research and development. Sufficient and sustained investment is a fundamental reason that the United States leads the world in science and technology. In this respect, China is still a 'developing' country.

As well as generally increasing investment, the Chinese government needs to target research areas with the potential to compete successfully in the international arena. This will help China to attract many of its overseas researchers back. Many Western-trained young scientists in the embryo biotechnology field

are returning to China, and they bring with them not only up-to-date expertise and technology, but also the Western mentality for teamwork and collaborative research. Now is the time for China to join the international competition for talent and to recruit and retain world-class scientists.

China does not have the resources to be competitive internationally in all fields of science and technology overnight. Its strategy of investing heavily in a few selected centres, as it did with the human and animal genome projects, should be applied to human-embryo research and related biotechnologies.

This approach is already being pursued elsewhere. In Australia, the government is providing US\$26.3 million to establish a Centre for Stem Cells and Tissue Repair at Monash University in Victoria. Even in the United States, where the federal government discourages this type of research, the University of California, San Francisco, announced a \$5-million developmental and stem-cell biology programme in 2002. In December of the same year, Stanford University revealed plans for an Institute for Cancer/Stem Cell Biology and Medicine and committed \$12 million in seed money¹². On the east coast, the University of Connecticut launched a Center for Regenerative Biology in 2002 with an investment of \$10.5 million to build the centre's labs plus additional support to recruit five new professors with complementary expertise to the existing strong embryo biotechnology programme. I had the privilege to be the founding director of this new centre.

Working together

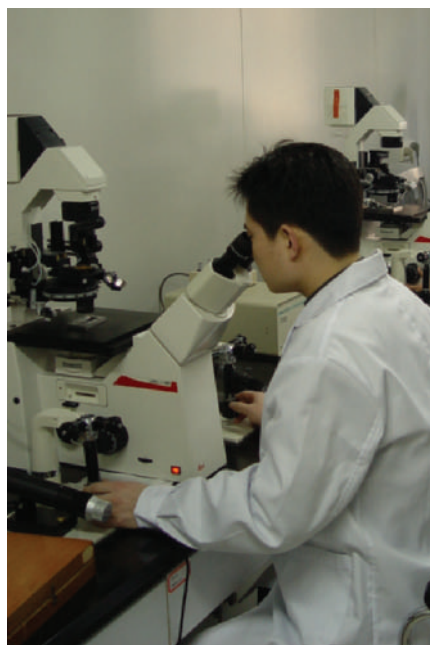
Whether or not China makes a clinical success of therapeutic cloning and other stem-cell therapies depends upon many factors. The gap between embryo research in China and other countries is not great at present. But, bringing these technologies through clinical trials and into the market will require expertise in a wide variety of disciplines, from molecular, cellular and developmental biology, to immunology and genetics, transplantation biology

Abuse of embryo technology will result in insurmountable problems for future research that no country in the world will want to face.

R&D budgets in different countries in 2000–01

Country	Total R&D (US\$ billion)	% GDP	Government R&D (% GDP)
China	12.5	1.1	0.28
United States	281	2.8	0.76
Japan	96.0	3.0	0.64
South Korea	20.0	2.7	0.64

Researcher at work in a cloning and stem-cell laboratory in Beijing.



and clinical medicine. The relatively undeveloped level of China's research base could pose major challenges at this stage. Cities such as Beijing, Shanghai and Guangzhou, with their solid foundation in basic science, should be considered as preferable locations to recruit national and international talents to establish research clusters and centres. These centres would consist of researchers with different expertise working together.

Another problem is the lack of collaborations both within and outside China. It is not uncommon in China for one university to have several similar labs studying the same area but with no collaborations between them; sometimes even unethical and unfair competition exists. In today's globalized world, no research should be done in isolation.

China has taken big steps to improve its national and international links, but there is still a long way to go, particularly when we look at pioneering research such as therapeutic cloning. Interactions such as collaborative research, conferences and joint publications between centres and other research laboratories should be encouraged by the government through funding and promotion incentives. These research centres in turn will attract more scientists to China.

An attractive proposition

Collaborations with China are becoming very attractive to researchers based in the West. While Western researchers focus on animal models, partners at the new Chinese stem-cell research centres could focus on human models. Furthermore, this will provide an opportunity for China to attract top scientists from around the world to work in China, full- or part-time, although salaries comparable to those in Western countries may need to be offered.

The world is beginning to take note of the Chinese researchers who are accomplishing excellent work. But we must realize that questions still remain in the minds of many Western researchers as to whether China has the ability to conduct world-class research. This is the result of long-standing inadequate communication and collaboration between Chinese scientists and those from the Western world.

'Opening the door', communicating

with foreign scientists, working together and jointly designing experiments, protocols and joint ownership of collaborative research data, patents and publications are principles and attitudes that many Chinese scientists still need to learn. Only by adopting and embracing such principles will the rest of the world get to know the best science and the first-rate scientists in China. I hope that these collaborative efforts will build a bridge of trust between this old Asian country and the West.

Guidelines

To conclude, I would like to emphasize the significance of governmental involvement, supervision and regulation of human therapeutic cloning and research on embryo biotechnologies. These techniques may involve human embryos, which reminds us that there are dangers as well as benefits: abuse of these technologies can result in insurmountable problems for future research that no country in the world wants to face. Research must be encouraged and supported, but clear guidelines are necessary for all researchers. A national ethics committee should be established to evaluate, approve, review and supervise therapeutic cloning and other embryo-based studies.

In Japan, any such study needs to be approved at the institute level by an internal

review board and by a central committee established by the government. This might provide a model for China to follow. Currently, the Ministry of Health is the regulating agency for human embryo and transgenic therapeutic research in China. But regulations are often not followed, and some very sensitive embryo-based studies are conducted with little or no institutional review, and researchers suffer no consequences for not following institutional or national regulations or guidelines, if they exist.

The government should create a national committee consisting of selected international experts in related fields to enforce these regulations. At the same time, local institutes should also set up their own ethics advisory/approval panels and be responsible for any violations. The national committee should also take responsibility for providing training for scientists and administrators on the current policies and protocols. This basic infrastructure should guide the emerging stem-cell therapies through clinical trials, and avoiding a rush into the clinic or even to the market before the basic science has been completed.

I expect to see research into stem cells increase in China. I would not be surprised if in five to ten years China becomes one of the leading nations in the world in the fields of therapeutic cloning and related research. But for that to happen, the Chinese government must provide strong support through policy, funding, talent recruitment and a determined commitment to promote productive scientific as well as business collaborations. ■

Xiangzhong Yang is the director of the Center for Regenerative Biology, University of Connecticut Storrs, Connecticut 06269-4243, USA.

1. Dennis, C. *Nature* **419**, 334–336 (2002).
2. Zeng, Y. *World Science and Technology Research and Development* (in Chinese) **5**, 1–4 (1999).
3. Wilmut, I. *et al. Nature* **385**, 810–813 (1997).
4. Ministry of Health, China. *Regulations on Assistant Reproduction Technologies and Sperm Bank*. www.moh.gov.cn/kjy/gzdt/wskj/1200309300005.htm (2003).
5. Zhou, Q. *et al. Science* **302**, 1179 (2003).
6. Chen, Y. *et al. Cell. Res.* **13**, 251–263 (2003).
7. Lu, C. *et al. Chinese Sci. Bull.* **48**, 1840–1843 (2003).
8. Hwang, W. S. *et al. Science* www.sciencemag.org/cgi/rapidpdf/1094515v1.pdf (2004).
9. www.chinanews.com.cn/2001-10-12/26/129578.html (2001).
10. OECD <http://international.tamu.edu/eucenter/Teich%20Presentation.pdf> (2002).
11. Office of Science and Technology Policy, USA. www.ostp.gov/html/R&DHighlights.pdf (2003).
12. Gershon, D. *Nature* **422**, 928–929 (2003).

A case for conservation

China urgently needs to take action to preserve its wealth of biodiversity, say Chung-I Wu, Suhua Shi and Ya-ping Zhang.

China is home to roughly 10% of the world's biodiversity. It has at least 30,000 species of vascular plants and the *Fauna Sinica* catalogue now stretches to 150 volumes. But this diversity is being destroyed by farming and industrial developments at an alarming rate. Conservation is therefore a matter of urgency, but to be taken seriously it needs a scientific basis and must attract the brightest scientists.

Remarkable achievements over the past five decades have been made in documenting flora and fauna in China. But it is now time to place an emphasis on understanding the general ecological and evolutionary principles behind the country's diversity. To achieve this, we believe that a national centre for ecological and evolutionary studies should be created.

Biodiversity basics

Before we examine this proposal, we need to understand the scientific basis of biodiversity research. The field divides readily into three areas: surveying and recording the flora and fauna; discovering how the species are related and distributed; and looking for patterns that help us to interpret this biodiversity.

Obviously, the third stage hinges on having sufficient information from the first two areas. After all, Darwin could deduce the general principles of evolution only after extensive studies had been carried out.

Scientists in China have made excellent progress in tackling the first two stages, and there are several ongoing mega-projects dealing with biogeography. For example, a team led by Deyuan Hong, a botanist based at the Chinese Academy of Sciences (CAS),



Mangrove trees are perfectly adapted to the harsh brackish environment in which they live.

Beijing, is examining the biodiversity in the Yangtze Basin, and ecologists Keping Ma and Ruyong Sun at CAS and Beijing Normal University, respectively, are studying bio-diversity conservation at key sites across China.

But what about the scientific principles underlying such studies? Biodiversity covers every aspect of ecology and evolution, from community stability to speciation (the duplication of species). Examples of the research areas it includes are: ecology and evolution; community and ecosystem; species competition and extinction; genetic diversity; formation of

species; and adaptation.

These areas highlight the variety of research in this field. Ecological and evolutionary principles are most easily discerned from simple systems. For example, grassland is usually a simpler ecological system than a tropical forest, and zebrafish are certainly easier to handle than tilapia (an African freshwater fish). The choice of a suitable system should take into account the ease with which general principles can be deduced. In China, complex systems were often chosen for economic rather than scientific reasons.

Basic research is also important for social

and cultural reasons. General scientific principles are extremely relevant to the urgent crisis in conservation biology. Looking at efforts outside the developed countries, it is clear that conservation has not been completely successful. The reason is not necessarily poverty alone, as this would imply that conservation ethics naturally follow the accumulation of wealth. China's wealth has grown extraordinarily, for example, but its conservation ethics have failed to keep pace. To close this gap, it is imperative that the country assembles a core group of biologists who are devoted to ecology, evolution and systematics.

Conservation biology therefore needs to be seen as a stimulating and prestigious scientific endeavour, otherwise few bright young people will enter the field. If the desired social status can be achieved, it will attract strong candidates to become college professors, school teachers and professional biologists. At the same time, these new recruits will be conservation-minded — as there are few ecologists, evolutionists and systematists who are not strong proponents of conservation biology at heart. From these professionals will spring the motivation for conservation.

A proposal for a national centre

More importantly, a national centre for ecology and evolution would help the drive towards a broader scientific understanding of biodiversity. Greater attention to fundamental science would encourage deeper and broader analysis and synthesis of biodiversity research. Ultimately, policy decisions will rest on such understanding.

At present, there are only a handful of institutes or departments in China where the emphasis is on the conceptual issues relating to ecological and evolutionary principles. Most noticeable is the lack of theoretical development. The stability of community structure, the population ecology of predator–prey (or host–parasite) interactions, the consequence of sexual rather than natural selection, the theory of speciation and the impact of population structure on adaptation are just a few examples of concepts that need to be developed.

To address this imbalance, we urge China's scientific community to adopt

strategies that will promote integrative research activities aimed at revealing general principles in ecology and evolution. At the local level, the formation of research and training programmes across traditional taxonomy-based departments (such as zoology and botany) should be strongly encouraged. But creating a new structure without imposing a cumbersome bureaucracy will be a major challenge.

A possible solution is to create an interdisciplinary training and research programme. This will allow scientists from existing departments to establish a new programme in, say, ecology and evolution, without creating a new physical entity. A more traditional approach is to realign existing departments to suit the new demand. Fudan University in Shanghai is likely to be the first in China to take the initiative and reorganize some departments to form a new department of ecology and evolution. This will probably be followed up by Sun Yat-Sen University in southern China.

At the national level, we propose the establishment of a national centre for ecological and evolutionary studies. This would promote the synthesis of the diverse array of biodiversity research results. The centre would not need permanent research staff, as its function would be to integrate data that already exist, rather than to create more. As a virtual institution, it would be administratively flexible and inexpensive to operate. We envision two main missions for the centre.

First, it would run workshops where ecologists, evolutionary biologists and systematists could gather. This would promote collaboration on joint analysis of results. The workshops would be significantly different from conventional academic meetings, where the role of the participants is to present and absorb. Instead, they would seek active participation in the analysis of all research results, wherever they are generated. The goal would be to integrate efforts across disciplines.

As part of this mission, the centre should also assume an international character by attracting participants from abroad. For example, in certain areas of evolutionary biology, particularly molecular evolution and population genetics, there is a strong contingent of Chinese expatriates who have

substantial commitments inside China.

In parallel with the workshops, a series of summer courses should be given. These would help to expose final-year undergraduates and postgraduate students to biodiversity research.

The second mission would be to train the next generation of ecologists, evolutionary biologists and systematists. The centre should act as a magnet for the brightest researchers in China. A number of PhD students would be sent by the national centre for sponsored research in the best institutions worldwide. The duration of training is likely to be about three years — roughly equivalent to doing a PhD.

Similar models

In many respects, the proposed centre would resemble the National Center for Ecological Analysis and Synthesis in Santa Barbara, California, and the planned US National Science Foundation Center for Synthesis in Biological Evolution. The scope and structure would be different, but the spirit of synthesizing existing observations into general principles would be the same. In fact, a prototype of the effort suggested here already exists in China in the form of the International Center for the Study of Evolution of Biodiversity, Kunming¹, which is at present partially supported by the Shanghai Institute of Advanced Studies, directed by Uli Schwarz. This institute is jointly sponsored by the Chinese Academy of Sciences and Max Planck Society, Germany.

The establishment of ecological and evolutionary research in China should prove beneficial for areas far beyond conservation biology. Many fields of biology in China will undoubtedly gain from enhanced efforts in ecology and evolution. These fields should be part of an increasingly integrative approach. After all, as Theodosius Dobzhansky said: "Nothing in biology makes sense except in the light of evolution." ■

Chung-I Wu is in the Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA.

Suhua Shi is in the State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China.

Ya-ping Zhang is in the Kunming Institute of Zoology, Chinese Academy of Sciences and Yunnan University, Kunming, Yunnan, China.

1. Normile, D. *Science* **302**, 559 (2003).

Agriculture of the future

Current technology will be insufficient to meet China's food demand in 2050. It is time to take action, says T. C. Tso.

China is experiencing strong economic growth. In 2002, its purchasing power parity (PPP), a useful indicator of economic standing, was second only to that of the United States. If China maintains its recent 5% annual PPP growth rate, and that of the United States stays at 2%, China's PPP will take the lead in 2023.

With an annual growth in gross domestic product (GDP) of 4%, China's per capita GDP in 2050 would equal that of Japan today and attain a level that is three-quarters that of present-day United States¹. China could surpass these levels if it addressed key social issues, such as the rural-urban divide, and harnesses the power of science and technology and its natural resources.

Disparities between rural and urban regions and between the east and west of China are all too apparent, and have resulted in a growing social gap. Currently, 70% of China's enormous population (predicted to reach a peak of 1.6 billion by 2030; ref. 2) lives in the countryside. Among this 70%, 50% are

farmers, and 20% are 'mobile' — they continuously migrate to areas where work is available. Those who don't move and instead choose to remain on their farm usually find themselves less than fully employed, because of the highly seasonal nature of their work. In addition, farmers' incomes are extremely low.

Reduce our burden

In the developed world, the agricultural sector accounts for a fairly low proportion of the GDP: in 2002, Japan's agricultural sector was only 1.4% of its GDP and 5% of employment, and in the United States it was 2% of GDP and 2% of employment. But in China, agriculture accounts for 15% of GDP and half of all employment^{2,3}.

Looking towards 2050, even assuming that the Chinese population stabilizes at 1.6 billion, demand for food is set to double. With the technology that is currently being used, it will be impossible to increase production to the level that is needed. China must act now to improve rural development.

The Chinese agricultural sector is far too large for any fast fix, but the Chinese government has for many years understood the importance of agriculture and has made solving agricultural problems a top priority. Despite this, basic problems remain.

To understand the agricultural situation in China, I often visit rural areas. I once asked a group of five farmers: "What message do you want me to deliver, on your behalf, to the president in Beijing?" They answered in unison, and without hesitation, "reduce our burden."

One pressure is agricultural tax, which contributes a mere 1% to the national coffers, yet is a heavy load for the country's farmers. This tax is in addition to costs for road repair, water systems and the maintenance of rural local offices, most of which is not paid for by the government. Although the fixed tax is low, the numerous other fees often mean that farmers pay out most of their income — and this can cause the occasional riot. Completely exempting farmers from national tax and local fees would not be a big financial loss to the government, but would lead to a huge increase in goodwill among farmers.

Another burden for farmers is the need to preserve land fertility and productivity, control erosion and reduce pollution. Understandably, farmers have other priorities: to obtain enough income they have to use the land intensively to grow multiple crops



M. HENLEY/PANOS

Challenge to China: how to develop its western provinces without creating industrial problems.

continuously, which results in more water, nutrients and soil being lost, as well as accumulation of salt from excessive irrigation. But the main reason for their lack of interest in these measures is that the land belongs to the collective. The farmers often feel like temporary caretakers, even though they may have long-term contracts. If the government granted ownership to the tillers, it would promote their sense of security and stewardship. Additionally, the many small and dispersed plots of each tenant could be redefined into sizes suitable for mechanization. This would lead to increased efficiency.

There are many other factors that could decrease the farmers' load and increase their income: application of modern technology, improving the credit and marketing systems, and, perhaps most importantly, reducing the surplus labour in the farming population.

Another way of increasing profits would be for farmers to organize themselves into cooperatives that encompass all steps from production to marketing. This would increase production efficiency, reduce waste, improve trading potential and achieve higher financial rewards because more money is made from processing than producing. This could be achieved by farmers' cooperatives working with organizations for rural enterprise jointly to develop expertise in agricultural operations, post-harvest management, and processing and marketing. Currently these enterprise organizations concentrate on industrial goods of low quality. Through this approach, the rural-enterprise organizations would contribute directly to increasing farmer's incomes and at the same time increase employment opportunities.

Resource development

To improve agricultural productivity and to take advantage of new initiatives, the three most important resources — people, land and water — must be used to their best effect.

Many would argue that the most important of these assets is the farmers themselves. But most of them have no opportunity to obtain a basic education, which makes the adoption of new technologies difficult. Children from farming communities need nine

Rivers in the south of China could be diverted to provide water for the north.



years of compulsory, free education (currently, education is not free and farmers generally cannot afford to pay the fees), and bright students from poor, remote regions should be helped to go to university, without regard for regional quotas; in the present system, rural students with the same academic qualifications are often denied higher education in favour of students from urban areas. The government and universities must promote rural education, either in the classroom or through distance-learning. An efficient extension system in the community should be developed to assist farmers in all phases of production and marketing, and to form a link between farmers and scientists. It should not be used as an organ for administering government policy.

China has devoted great effort to developing its science and technology base, and has made remarkable progress. The objectives of this policy are to improve general welfare, increase national and international competitiveness and promote academic achievement. But, as measured by indicators such as the number of biotechnology and chemical patents and the innovation index, China still lags far behind other countries. It must increase its science and technology investment, which will ultimately lead to poverty reduction and to increased agricultural production.

According to a recent report⁴ by the International Food Policy Research Institute in Washington, every 10,000 yuan (US\$1,250) that the government invests in research and development brings seven people out of poverty. In agricultural production, every yuan invested in research yields 8.3 yuans in increased productivity.

China can benefit from the experiences of developed countries in restructuring its science and technology. In agriculture pro-

grammes, education, research and extension must work as one unit. Each province or region should have its own, nationally coordinated, higher education system similar to that of the US land-grant institutions, which receive federal support for integrated programmes of agricultural teaching, research and extension. Some regions, because of their large size, may need several such systems.

To feed China's population in 2050, a scientifically educated workforce with an agricultural background will be vital. China may well need to use low-risk, genetically modified organisms and transgenic animals² to meet its needs — and scientific knowledge will be essential to ensure that they are successfully implemented.

Many scientists in China are attempting to increase plant yields and resistance of plants to stress. However, a lack of nationwide coordination and cooperation has resulted in duplication of effort and low efficiency. China has recently established many pioneer laboratories for basic and applied research, but a national strategy is still needed to ensure that limited human and financial resources are not wasted.

As well as education, farmers need land to cultivate — and yet, despite China's size, this is a resource that seems to be scarce. Only 10.2% of China is suitable for agriculture, and 37.1% is suitable for animal husbandry. And the amount of arable land is set to shrink by one-fifth because of projects to construct and restore pastureland and forest^{2,3}.

There is great concern in China about achieving agricultural self-sufficiency, although to some extent this worry may be misplaced. It is almost impossible for any





The California of China? Western provinces that are rich in resources should be developed.

resources, and so a more viable alternative may be to use the vast supply of water from the west for the north and east. There are two ways in which this could be done. One is to divert water from the upper reaches of the Jinsha, Lancang, Nu and Yarlung Zangbo rivers (see map). Another way is to divert water from the 'big U-turn' of the Yarlung Zangbo in Tibet. This river has an annual run-off of 165 billion cubic metres, and the 'narrow neck' across the U-turn is only 40 km, but has a drop of 2,250 metres. This stretch has an annual flow discharge of 1,900 cubic metres per second and an annual run-off of 60 billion cubic metres. A 16-km tunnel could be built near the neck, connecting it to local waterways. The natural hydropower from the drop would generate sufficient energy to transfer the water to the north and northwest. If it was then connected to the upper reaches of the Jinsha, Nancay and Nu rivers, the total capacity would meet China's needs in the future.

For China to produce enough food to feed its population at the middle of this century, it needs to use its precious resources wisely — be they human, land or water. The path from the green revolution (the improvement of agriculture by using modern farming techniques such as high-yield plant varieties, irrigation, adding fertilizer and pest control measures) to the gene revolution must be navigated step by step. The recent increase in funding of scientific and agricultural research is very welcome, but to solve all its agricultural problems, China must allocate more funds, up to 3% of GDP (from its present 0.9% level). If it does this, China's future will be very healthy and its excellent recent annual growth will be sustained. China could then take its place amongst the most developed nations, such as Japan and the United States. ■

T. C. Tso is a retired senior scientist, currently working as a collaborator of the national program staff at the agricultural research service, US Department of Agriculture, Beltsville, Maryland 20705, USA.

1. World Bank's World Development Indicators Database, July, 2003.
2. Lampe, K. *et al.* *Vision of 2050 Agriculture in China* (ed. T. C. Tso) (IDEALS, Beltsville, Maryland, in the press).
3. Tso, T. C. *et al.* *Agriculture in China 1949–2030* (IDEALS, Beltsville, Maryland, 1998).
4. Tso, T. C. *et al.* *Vision of 2050 Agriculture in China* (ed. T. C. Tso) (IDEALS, Beltsville, Maryland, in the press).
5. Fan, S. G., Zhang, L. X. & Zhang, Z. B. *Growth, Inequality, and Poverty in Rural China: The role of Public Investment*. IFPRI Research Report 125, International Food Policy Research Institute, Washington, DC, 2002.

one country to satisfy all of its own needs — every country is limited by its local conditions and must therefore rely on trade to obtain what it cannot produce.

In China, there is considerable scope for agricultural expansion, if previously unusable land can be developed, perhaps through enhanced production systems, better irrigation and new technology. For example, there are 8 million hectares of arable land and 47 million hectares of pastureland that are as yet undeveloped in Xinjiang region alone, and there are many more areas ripe for development in other western Chinese regions.

Migration to those areas could ease population pressures in the east of China, where 74% of the population lives, while increasing China's overall agricultural production. Although the world generally takes a dim view of such large-scale migration because it destroys cultures, it is needed for China to improve the quality of life in this vast rural region for original westerners and new settlers. The aim of this 'China west' development would firstly be to produce enough food, economic and social opportunity for this area of China and secondly to address the needs of the east. Among the vast western regions, Xinjiang, because of its abundant resources, should enjoy top priority in development, and this will ensure the fastest returns on investment. Xinjiang could become the 'California' of China³.

An alternative is to 'rent' arable lands abroad, in countries such as South America, Africa and Russia, and for Chinese farmers to operate there under contract to meet Chinese needs^{2,3}. But the long-term solution still rests with China itself, and will require care-

ful land management as well as science and technological development.

Water is the other precious resource for agriculture. China's supply of underground water is limited, and surface water is unevenly distributed, with only 20% of water collecting in the areas where 64% of arable land is located. By 2050, China's total water deficit could reach 400 billion cubic metres (roughly 80% of current capacity). China's present water supply offers only slightly more than 500 billion cubic metres, which is below current demand, and is estimated to be 10–15% higher than the quantity available³.

No rain on the plain

Rainfall distribution is uneven, and the northern plain has more than half of the arable land, but only 20% of rainfall. Most rainfall is in the southeast regions. In the west there is plenty of water from high mountains, which are the sources of the Yangtze and Yellow rivers. But most of these western waters run through the desert or to foreign countries and cause flooding.

The government has addressed this problem by various means, but a solution has proved elusive. One answer would be to fully use all water resources by conservation and diversion.

The current plan for a northern diversion of the Yangtze would not benefit agriculture, however. Furthermore, it is questionable how much water from the Yangtze can be diverted considering that 40% must be retained to maintain local ecology and 20% is needed for additional economic and social growth in the Yangtze region.

The west is rich in mineral and water

Making big money from small technology

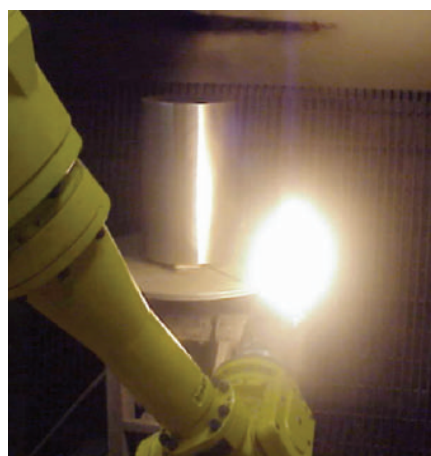
With venture-capital funds depressed, kick-starting a technology business can prove to be problematic. James C. Hsiao and Kenneth Fong offer some advice for budding entrepreneurs.

China has invested heavily in technology in recent years to lift its rapidly growing economy to a dominant position on the world stage. There is no doubt that the country is making its presence felt in the emerging fields of nanotechnology and biotechnology. But converting research into profitable products is tricky. Examining the ingredients of a successful company and looking at alternative approaches to funding should help to transform research ideas into profits.

The development of nanotechnology in China began with China's ten-year 'climbing up' project, which supported nanomaterial research between 1990 and 1999 (ref. 1). This project was very much ahead of its time — and was started many years before President Bill Clinton created the National Nanotechnology Initiative in the United States in 2000. US federal nanotechnology research funding has since increased by nearly seven times, to a budget of US\$847 million in 2004 (ref. 2). China's investment level is comparable.

There are several sources of funding for new technology in China, including the National High Tech R&D Program of China, the National Natural Science Foundation, the Ministry of Science and Technology and the Chinese Academy of Science³.

The high number of papers at recent nanotechnology conferences in China provides evidence of the vast amount of research and development that is being carried out by Chinese researchers, and shows that significant commercialization efforts are being focused around Shanghai and Beijing. The establishment of the Nano Science and Nanotechnology Center of the Chinese Academy of Science and the Shanghai Nanotechnology Promotion Center are good



A robotic arm holds the plasma torch for manufacturing ceramic nanocoatings.

indicators of the level of interest.

But what does it take to create a profitable technology business? The key principles for growing an emerging technology company are basically the same anywhere in the world — and they apply to any technology.

The first step is obvious: the new technology must work. The next steps are probably less clear but are just as important. You need to recruit a good management team and to demonstrate that consumers will want your products. This is known as 'market pull'.

Scientists and inventors tend to overestimate the value of their invention and often fail to consider whether the problem the product solves is significant enough to capitalize on market pull. As a result, they get into a situation where instead they find themselves having to 'push' their technology, a sure recipe for failure. This is why venture-capital (VC) companies invest in people rather than technology.

Badly burned by huge losses from the burst of the dotcom bubble in 2001, the VC

world is playing safe and is 'keeping its powder dry'. VC investment is now mostly limited to follow-on funding to protect past investments, and is generally characterized by investment rounds in which valuations are either unchanged or reduced. Moreover, VC firms now try to hedge their bets by using complicated financing deals.

As a result, investment contracts may specify a liquidation-preference clause, in which the VC company gets a predetermined rate of return (such as twice the level of investment) in the event of sale or liquidation of the company. The investors get their money back, perhaps many-fold, before the entrepreneur sees a penny.

Another way for VCs to protect their investment is to insist on preferred shares, which give investors a fixed dividend from the company's earnings and entitle them to be paid before common shareholders.

VC funding is usually handed out in stages. One way is milestone financing, where an investor provides additional funding as long as certain conditions, or milestones, are met. Steps such as prototype development, obtaining intellectual-property protection and beta-testing of a product may be viewed as milestones.

Wealthy individual investors, known as 'angels', have become the real VC funders in today's investment scene. These investors tend to be more interested in helping start-up companies achieve organic growth (which excludes any growth from takeovers, acquisitions or mergers). Angel investors are usually in for the long haul, and terms are generally much less onerous than with VC firms.

Nevertheless, given the present funding climate, emerging technology companies are clambering for alternative approaches to fuel their growth and technology development. But before any company can think of funding, it needs to be sure of its intellectual-property rights.

Badly burned by losses from the burst of the dotcom bubble, venture capitalists are now playing safe and 'keeping their powder dry'.

Both nanotechnology and biotechnology are receiving increased publicity globally. Marked advances in both fields are the result of years of work by scientists and millions of dollars of research funding. In addition, both nanotechnology and biotechnology tend to be highly interdisciplinary, involving scientists from widely differing backgrounds and with different skills. Intellectual property is driving the nanotechnology and biotechnology revolutions, and patent portfolios are the currency.

For instance, patent offices are now flooded with applications that are relevant to nanotechnology — and the Chinese patent office is no exception. Recent enforcement of intellectual-property rights since China joined the World Trade Organization has led to greater respect for patents in China and has improved the country's investment climate. For example, Tsinghua University Nano Center has received funds from Hon Hai, one of Taiwan's largest technology companies.

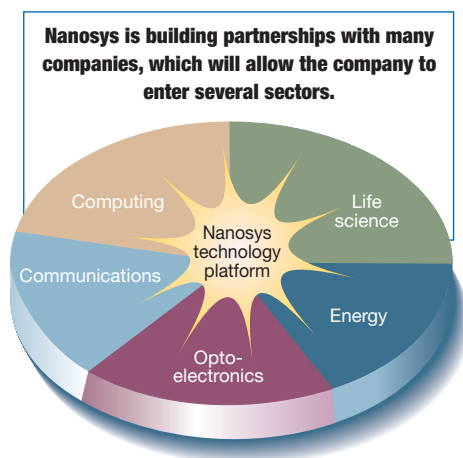
Materials science

Nanotechnology can be divided into two groups — materials and devices. Historically, venture capitalists have shied away from investing in materials, largely because of slow development times as a result of large companies having to check the performance of the material in their products. This can lead to poor return on investment, which is often exacerbated by inevitable commoditization — when the material reaches the mainstream market and other manufacturers can copy it.

But the lengthy product lifetimes of materials can make them very profitable over the long term⁴. Moreover, materials are 'disruptive', in other words they lead to new products that can replace present technology, allowing for a vast number of product opportunities in the mainstream as pricing drops because of increased production⁵.

By contrast, VCs love to invest in devices, where margins are high, although product lifetimes may be short. Of course, in the nanotechnology sector, devices are still years from commercialization. The same is

Intellectual property is driving the nanotechnology and biotechnology revolutions, and patent portfolios are the currency.



largely true

for biotechnology. Nevertheless, VC firms such as Lux Capital (New York City) and Ardesta (Ann Arbor, Michigan) are specializing in nanotechnology and are making early-stage investments.

Given the difficulty in obtaining funding, particularly for materials science, start-up companies are searching for alternative sources of cash. There are three main routes that they can pursue: government contract research and development; corporate partnering through joint-development programmes; and corporate investment. As an illustration of these approaches, this article highlights the paths taken by three companies.

Nanosys, a nanotechnology company based in Palo Alto, California, has taken advantage of corporate investment opportunities thanks to its executive chairman and co-founder, Larry Bock, who has extensive experience of setting up companies. By contrast, nanomaterials firm Inframat, in Farmington, Connecticut, has used government contract research and development funding to grow organically. And Epiomics, a biotech company in Burlingame, California, has expanded by establishing corporate partnerships, following two rounds of seed financing.

Nanosys was founded in 2001 and has since benefited from the experience of Bock, a serial entrepreneur who has an excellent track record in building successful companies. Of his 12 previous start-ups, two are privately held, two were acquired, and eight

now trade on the US high-technology stock market Nasdaq. So far Nanosys has raised in excess of US\$70 million — and it has done this through investor confidence in a management and scientific team that has proved it can deliver.

The company amply demonstrates the maxim that VC firms invest in people, not technology, as Nanosys does not expect to market any products based on its core technology for a few years, although it expects to deliver some niche-market products soon to generate revenue. But it has aggressively licensed intellectual property from a variety of sources, including Harvard University, the Massachusetts Institute of Technology, the Lawrence Berkeley National Laboratory, University of California, Berkeley, and the Hebrew University of Jerusalem, to attain dominance in the field of inorganic semiconductor nanostructures. The result is that it now holds over 140 patents). Applications for this technology include solar panels, biosensors and flexible electronics — semiconductors fabricated on plastics, for example. Interestingly, the latest round of investment in Nanosys featured funds from three Asian sources: the China Development Industrial Bank, Chiao Tung Bank and United Overseas Bank in Singapore.

Building a nest egg

The business strategy pursued by Nanosys is to commercialize technology through partnerships with other companies. The pie chart shows the range of industries with which the company hopes to set up links.

Entrepreneurs tend to think in terms of the minimum amount of funding needed to accomplish their goals. This is a mistake. There is no doubt that seasoned entrepreneurs such as Bock focus instead on building a considerable 'nest egg' of funding to take the company forwards, concentrating on valuation, rather than getting hung up on dilution (a decrease in equity). At the end of the day, it is better for the entrepreneur to maximize the value of his or her own equity position, even if that means giving away more of the company to achieve a higher overall value.

For its part, Inframat has used a non-dilutive bootstrap approach, which is when organic growth is achieved in an incremental fashion — as money comes in, some

progress is made, which then attracts more money and more progress. Inframat's approach has been based on US government contract research and development, in which the contract funding has been used to achieve organic growth without the help of outside private capital. From an investment angle, we can call this customer-based financing, and in this case the customer is the government.

Founded in 1996, Inframat has pursued opportunities in the Small Business Innovation Research Program and the National Institute of Standards and Technology's Advanced Technology Program. Inframat's core technologies are based on using aqueous chemistry to synthesize nanomaterials cheaply, and on a thermal spray that is used to produce nanostructured coatings. For example, the company has developed dense ceramic nanocoatings that can be used to prevent sea creatures from adhering to the metal bodies of ships.

As with Nanosys, forming partnerships has been essential for Inframat's growth. An early association with the University of Connecticut was key to the company winning large government contracts from the US Navy. Over \$5 million in funds has resulted in nanocoatings for submarines, aircraft carriers and minesweepers.

Inframat's most important technology development is its thermal-barrier coatings for gas turbines (both flight- and land-based). This technology improves the lifetime of the turbine, and Inframat is now approaching land-based turbine manufacturers and refurbishers.

This method of gaining funding from the government has been recently adopted in China. The country now has its own small-business innovation research programme, called the Innovation Fund for Small Technology-Based Firms. The first handbook for applying for funds appeared in 2002. This funding should focus on early-stage technology development, where the benefit to mankind is potentially high, and where the risk is perceived to be too high for corporate investors.

Epitomics took a different funding path from Nanosys and Inframat, but once again it focused on creating strategic partnerships



The handbook for applying to China's Innovation Fund for Small Technology-Based Firms.

with large firms. The science of epitomics is new. It is the study of all epitopes of the proteome in an organism. The epitope is a functional recognition site on a protein that can be bound by a specific monoclonal antibody. Epitomics is an emerging company that is developing specific antibody technology that was first licensed from Loyola University in Chicago and the University of California, San Francisco.

The company aims to become the leader in rabbit monoclonal-antibody technology with products for research, diagnostics and therapeutics. It has a wholly owned subsidiary in Hanzhou, China, and will probably spin off a subsidiary to develop its own portfolio of products for the diagnostic market. A similar strategy will apply to the therapeutic monoclonal-antibody market.

The key strategy for growth that Epitomics has used is entering into service agreements with private and public biotechnology and pharmaceutical companies. At least ten companies have now signed up, including Genentech and SUGEN in South San Francisco, California; Cell Signaling Technology in Beverly, Massachusetts; Exelixis in San Francisco; and Upstate USA in Waltham, Massachusetts.

A successful business requires an excellent intellectual-property portfolio, proof of market demand and a management team that can deliver on its promises.

Clearly, this approach will enhance valuation because the company's technology has been shown to be useful — and the generation of sales revenues is always a bonus.

The main point here is that the partners generate channels of distribution for the products, which saves on marketing. Epitomics has also completed a recent Series B (second) equity round, with investment from Kenson Ventures in Menlo Park, California, and Bill Rutter (founder of biotech firm Chiron). Just as with government funding, rapid-growth companies also need to raise some private equity.

In summary, the fundamental attributes of a successful nanotechnology or biotechnology strategy in the United States and China should be based on an excellent intellectual-property portfolio, proof of market demand and a management team that can deliver on its promises. As can be seen for all the profiled companies, building partnerships with government agencies and other companies is another important route to success.

Exceptional synergies between the Chinese and US economies are evolving. China's open-door policy, which encourages US investment in China's technology sector, will no doubt lead to the active involvement of US and other Western companies in future start-ups in China. China will follow all three models presented here for starting and growing a company in the nanotechnology and biotechnology sectors. As a result, we believe that China will lead the world in both fields. No question. ■

James C. Hsiao is co-founder and chairman of Inframat, 74 Batterson Park Road, Farmington, Connecticut 06032, USA.

Kenneth Fong is chairman of Kenson Ventures, 695 Oak Grove Avenue, Menlo Park, California 94025, USA.

1. Bai, C. J. *Nanoparticle Res.* **3**, 251–256 (2001).
2. Paull, R., Wolfe, J., Hebert, P. & Sinkula, M. *Nature Biotechnol.* **21**, 1144–1147 (2003).
3. Hsiao, J. C., Ma, X., Xiao, T. D. & Reisner, D. E. Thinking Small in China – Is Nanotech the Next Big Thing? in *Proc. 5th Annu. BCC Conf. Nanoparticles 2002* (Business Company Communications, New York, 2002).
4. Moore, G. A. *Living on the Fault Line* (Harper Business, New York, 2002).
5. Christensen, C. M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* (Harvard Business School Press, Boston, 1997).

Acknowledgements: The authors thank Larry Bock of Nanosys for allowing use of the pie chart, and David Reisner of Inframat and US Nanocorp for helpful discussions.



Follow your nose

Alice Shih-hou Huang draws on her own experience to highlight the many careers and opportunities open to scientists in the West and in China.

When I visited Beijing a few years ago, I asked the young people I was talking to to name a twentieth-century hero. They did not name a politician, movie star or a millionaire; they named the scientist Albert Einstein.

Scientific research is now considered one of the most prestigious occupations in China — with a reputation to rival that of the ancient imperial scholars. Historically, parents identified their brightest child and prepared them for intellectual pursuits by hiring tutors. After successfully completing a series of national exams, the brightest of them were included in the emperor's court as valued advisers. Such exalted levels were only achieved after many lonely hours spent studying and practising calligraphy.

Science is now attracting the same calibre of students, but those advocating this discipline often know little about modern research. Gone are the days of rote learning, now scientists are equipped with a wide range of skills, from problem solving to communication. These accomplishments make students and researchers ideal for many careers. If you're brave and follow your instincts, there are amazing opportunities open to young scientists — in China and around the world.

Before offering advice to young people wishing to become scientists, we should ask why this area is so attractive. Parents encourage their children into this field because it is a respected profession, with low unemployment. Students, thinking of well-known scientists, expect fame and fortune, and teachers direct students into science when they see pupils with rational and quantitative capabilities, believing that these attributes will be rewarded.

New skills

But those advocating science still often see it as a technical trade that requires students to memorize endless facts. In reality, there is too much information for such an approach to be useful. Also, problems that involve massive quantities of data are much easier and more accurately solved by computer than by a researcher laboriously working through the information.

Technical expertise and 'good hands' are still valued in the experimental sciences, but training in scientific thinking, priority setting, problem solving and clear communication is just as important for success in the international scientific arena.

There is no longer a rigid path to success.

To face this new world, scientists have to be flexible during their training and throughout their careers. They must be open to new opportunities and be willing to move to where the best jobs are.

When I started studying medicine, I hardly knew what to expect. In fact, I could not have imagined the life I am leading now. Ever since I was in Kweiyang (southwest China, Kweichow province) as a young child, I admired the missionary doctors and thought that the most wonderful thing I could do when I grew up would be to cure diseases and save lives.

When I had the chance to enter Johns Hopkins School of Medicine in Baltimore, Maryland, I grabbed it. But it wasn't very long until sick patients with bedsores and the time-consuming, repetitive boredom of dealing with common illnesses took away any romantic notions I had about medicine. Research to uncover the causes of illnesses and find new ways to combat them seemed much more exciting. At the time, new PhD programmes aimed at medical students were being developed, and I jumped at this chance.

Upon entering the programme and carrying out serious research for the first time, I knew I had found something I was better suited to, and that I loved. Spending the rest

Science in China is often seen as a technical trade requiring students to memorize endless facts.

of my days in a lab working on viruses would have fulfilled my dreams.

I grew viruses, purified them and counted them by observing their indirect effects on cells grown in a single layer in a Petri dish. In one of my first experiments, in which I was purifying viruses, I discovered a separate subclass of viral particle that was defective and could interfere with the growth of the standard virus¹. This was very exciting: a mutant particle such as this could be used to control viral diseases, especially those of plants. Its properties so intrigued me that I did not finish the requirements for the MD degree after I received my PhD.

One day, the head of the lab challenged me to become a professor. This thought had never crossed my mind, because I saw so few women professors, although there were many female researchers. This was a new goal.

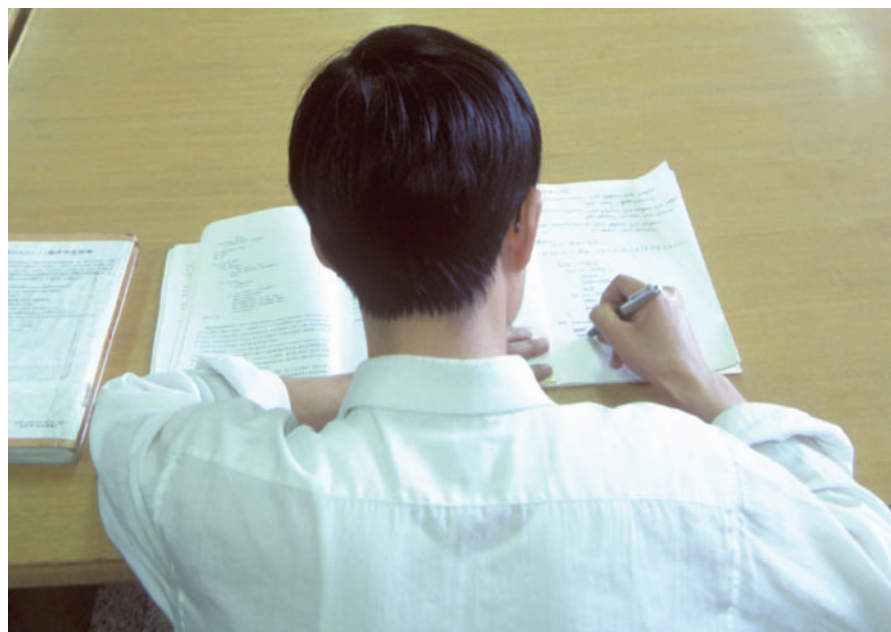
I continued to enjoy the research, but I took on extra responsibilities: teaching, being a member of various university committees, leading a division studying infectious diseases, and providing advice to numerous non-profit organizations and government agencies. After a while, I was promoted to a tenured professorship at Harvard Medical School, but as much as I enjoyed it, I wondered what to do next.

The dean advised me that being a professor is the perfect job, with considerable freedom and power, especially in the United States. Professors often keep working and maintain their own lab and students until they are well into their 80s, so long as they obtain independent funding from government or non-profit agencies.

But the long scholarly life was not for me. I was sure that I could continue to produce good research, but I also realized that the money I obtained for my lab would probably be better used by a younger investigator with fresher ideas. I looked around for an alternative and was attracted by research administration.

When I moved from Boston to New York, an opportunity to become dean of science at New York University was offered to me. Although I was supported by a Merit Award from the National Institutes of Health at the time, I was able to explore the new administrative role together with my research for a three-year period.

Now as senior councilor for external relations, a part-time position, I coordinate different areas of science and promote ideas at



the frontier of more than one discipline. Also, I vicariously enjoy the success of the next generation of scientists whom I am able to help. To nominate worthy colleagues for promotion or awards is another gratifying experience.

The number of consultations both within the university and externally continues to increase. They come from institutions as diverse as universities, foreign governments and investment groups. In addition, I receive many invitations to give talks on career development, science policy and diversity from professional groups. Unlike research, which is a very competitive and relentless, my activities take place at a more leisurely pace, so that there is time for family interactions, travel and new hobbies.

Instinctively right

It is not unusual for scientists these days to follow different career paths. Many go into academic administration or into various biotechnology or pharmaceutical companies. An extreme alternative is going into non-profit advocacy: an example being the molecular biologist, Michael Jacobson, who is the founding director of the Center for Science in the Public Interest, based in Washington DC, which protects the consumer by examining prepared foods to ensure that their ingredients are properly identified and healthy.

Other scientists, especially in developing

countries, have gone into politics and, either by election or appointment, become high-ranking government officials.

My parents were unhappy when I did not complete my initial training as a doctor, but by following my nose and knowing my own strengths, I made the right decision for me.

To those readers who are more established in their careers and responsible for the education of scientists, I say that it is imperative that the training their students receive should be more than just technical and limited to a narrow field of science. Instead, they should develop the ability to choose an important problem, ask the right questions and design hypotheses—all essential tools for scientists. Young scientists should be encouraged to explore new paths and to have the confidence to follow their instincts.

If you are an aspiring young scientist in China, you are entering a wonderful career, full of all sorts of possibilities. You may end up in academia, but you may also end up working in areas such as financial investment, journalism or philanthropy. Do not limit yourself. A science background prepares you for all these possibilities. Do not bind your feet to prevent your own progress.

Alice Shih-hou Huang is senior councilor for external relations and faculty associate in biology at the California Institute of Technology, Mail Code 1-9, Pasadena, California 91125, USA. ■

1. Huang, A. S. & Baltimore, D. *Nature* **226**, 325–327 (1970).

A question of biology

Complete genome sequences have provided a plethora of potential drug targets. But the hard task of finding their weak spots is just beginning, as Caitlin Smith finds out.

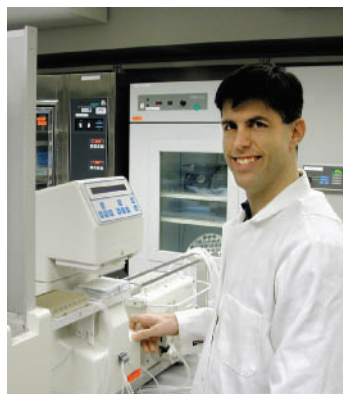
With the complete sequencing of the human genome, one might think that the identification of new potential drug targets would be coming to a halt. Yet there are still many exciting developments to come in the field of target identification, with technological advances enabling challenging biological questions to be addressed in increasingly creative ways.

According to Kevin Fitzgerald, group leader of applied genomics at Bristol-Myers Squibb headquartered in New York, one of the biggest challenges is to understand the underlying physiology of drug targets. "We do not yet understand all of the basic underlying principles and rules," he says. "We are left trying to interpret and predict the behaviour of dynamic processes based on small snapshots in time. The electrifying part of being involved in biology today is that the technology is beginning to catch up. If you are able to take enough snapshots under different stimuli and conditions, over time, concepts and rules do emerge."

Part of this challenge is to identify interactions between drug targets, which are generally proteins such as receptors and

enzymes, and the proteins and other molecules that regulate them. As Jan Mous, president and chief executive of IntegraGen of Evry, France, says: "The most exciting development is also the most challenging: to first understand and appreciate the subtle interaction of many proteins and different messengers and hormones steering the function of our cells and organs, and then pick the best target for drug interaction giving the most favourable ratio of wanted versus unwanted effects."

Interactions are often subtle and complex, requiring a multidisciplinary approach and the use of informatics systems to make sense of the increasingly complicated data. "A very tight integration of computer sciences and mathematics with experimental biology and chemistry has to be the modus operandi for the future drug-discovery business," says Mous.



Kevin Fitzgerald: finding rules.

One approach to investigating physiology is through genomics. Now that the human, mouse and rat genomes have been sequenced, high-throughput genomics can come into its own. Ingenium Pharmaceuticals in Martinsried, Germany, uses large-scale chemical mutagenesis and screening to look for novel, medically relevant biological

mechanisms in mouse and rat mutants. When a phenotype of interest is found, the mutation is located by high-speed positional cloning and a line of model animals can be produced. Further biological information can be gathered through gene-expression and cell-pathway analysis, complementing the pathophysiology observed in the mutant animal.

A new mechanism of sensitization to the hormones leptin and insulin was recently discovered using this technology. "The effect

P. KATNE

RUNNING INTERFERENCE ON THE GENOME

RNA interference (RNAi) is a powerful method of gene silencing using double-stranded RNA, which is converted to small inhibitory RNA (siRNA) as the active agent. siRNAs are now hot property as reagents for knocking down gene expression in target identification and validation assays, and as possible therapeutics for silencing genes associated with disease.

According to Christophe Echeverri, chief executive and scientific officer of Cenix BioScience in Dresden, Germany, high-throughput RNAi is the most exciting recent development in target discovery and validation. "The key advance here is the development of genome-wide libraries of siRNAs to enable genome-wide screens in human cells," he says. "It offers the most direct, cost-efficient way to get from gene sequence to gene function at a genomic scale." As part of its drug-discovery and target-validation programmes, Cenix has designed siRNAs against more than 98% of the human, mouse and rat genes identified by the genome-sequencing projects; the siRNAs themselves, for all these genes listed in the RefSeq database, are available to order through Ambion of Austin, Texas, which also offers an off-the-shelf



Christophe Echeverri thinks big with siRNAs.

library of nearly 600 kinase siRNAs made to the Cenix designs.

The siRNA drug-development company Atugen in Berlin, Germany, also develops siRNAs as both reagents for target identification and validation and as potential therapeutics. Atugen's technology platform for target identification uses various messenger RNA knockdown techniques, including proprietary siRNAs chemically modified for increased stability, and a proprietary antisense method called Genebloks. "These are used in combination with proprietary liposomal formulations to elucidate gene function in specialized cellular assays and animal models," says Klaus Giese, chief scientific officer at Atugen. "Matching results found with siRNA and Genebloks will substantially increase the probability that an observed phenotype is clinically relevant."

RNAi allows more questions to be asked of mammalian cells than ever before. "The ability of RNAi to allow the 'yeastification' of mammalian cells in terms of genetics is very exciting," says Kevin Fitzgerald, group leader of applied genomics at Bristol-Myers Squibb in New York. "With RNAi we can now begin to approach the genetics of tumour-suppressor genes in a whole new way. We have known that the loss of these genes is very important in the development of cancer for years, but we have been unable to develop specific drugs directed at them. With RNAi we can now ask whether there are any druggable targets in the genome that will lead to the death of cells missing tumour suppressors but leave normal cells alone. This is a question that in mammalian cells we simply could not have asked a few years ago."

C.S.

of the inhibition of the newly discovered pathway is that our mouse model has highly improved tolerance to glucose while requiring only a fraction of insulin for this regulation," says Ingenium's chief executive Michael Nehls. "In contrast to other insulin-sensitizing effects, the mice do not become obese — on the contrary, they slim under a high-fat diet. We believe this could be a breakthrough in obesity/diabetes research," says Klaus Dembowski, vice-president of drug discovery

From genotype to phenotype

Despite the increasing emphasis on proteomics in target identification, DNA microarray technology is still a powerful technique for identifying genes involved in susceptibility to diseases. Target discovery and identification should soon be benefiting from the DNA microarrays of the complete human genome now sold by several companies. First to market in July 2003 was NimbleGen in Madison, Wisconsin, with a chip containing 200,000 probes, with an average of five probes per gene. This was followed in October 2003 by the Human Genome U133 Plus 2.0 array from Affymetrix in Santa Clara, California, with 1.3 million probes and the ability to analyse the expression of about 47,000 different transcripts. The most recent human whole-genome chip, launched last month, comes from Agilent Technologies, in Palo Alto, California. Agilent's double-density-format chip represents about 41,000 genes and, with Agilent's ImaGene image-analysis software, is compatible with most commercial microarray scanners for 25 × 75-mm chips.



What it means to be human?

This open-platform approach, the company claims, will make it easier for scientists to move from their own homemade chips to the Agilent chip.

Microarrays are already widely used to study gene expression and to detect sequence features such as single-nucleotide polymorphisms (see *Nature* **422**, 917–923; 2003). Future uses will include assays to study the many forms that a protein target may take as a result of alternative splicing. "For RNA analysis, over the next year, we will start to see

assays for efficient splice variant analysis that will enable us to move from gene-specific analysis to transcript-specific analysis," says David Craford, vice-president of business development at Affymetrix. "More data are emerging to suggest that for many diseases this will be incredibly useful."

According to Greg Yap, senior marketing director in DNA analysis at Affymetrix, one of the newest and most exciting applications of microarrays to target identification is the possibility of genome-wide association studies of complex genetic diseases. "The experiments that geneticists really want to do are to find genes that cause complex diseases in unrelated clinical populations," says Yap. "Finding genes that cause a disease in a family makes a nice story, but finding genes causing a disease in the world makes a drug target." Studying genetically unrelated people is typically difficult, owing to the numbers of genetic markers per person that need to be studied. The advent of microarrays has made individual genotyping feasible, "making it realistic to compare the genomes of people with and without a disease at high-enough resolution to find potential drug targets", he says.

Pharmaceutical company Roche in Basel, Switzerland, is using the Affymetrix whole-genome expression arrays to identify genes that play key roles in disease pathways. "One approach for new target discovery that we use is to identify the genetic basis for disease susceptibility using a coupled genetic and genomic analysis," says Gary Peltz, head of genetics and genomics at Roche. "Having the complete genomes of several organisms (human, mouse and rat) available on chips is a major advance. It allows complex biology,

AFFYMETRIX

DRUG DISCOVERY IN REVERSE

Reversing the current drug-development path of 'target to drug to phenotype', chemogenomics, or chemical genomics, starts with a known drug that causes an interesting disease-relevant phenotype *in vitro* or *in vivo*, and then identifies the cellular target(s) for that drug. Chemogenomics is a powerful tool in target identification because you start with an active small molecule. By definition, any targets found are 'druggable' and are likely to be significant in the phenotype caused by the drug.

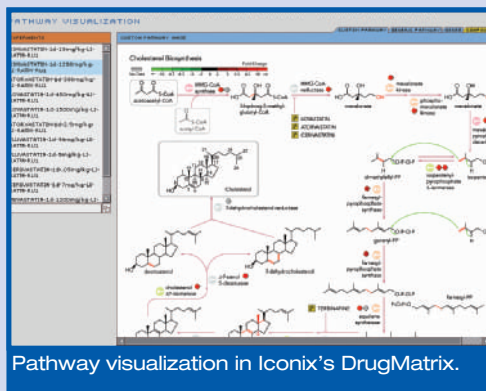
Seth Cohen, who recently moved from Millennium Pharmaceuticals to Caliper Life Sciences in Mountain View, California, likens chemogenomics to "reverse drug discovery". In other words, you take a drug compound with known effect and analyse in detail the result of exposure of the cell or organism to the compound. "This allows for a more precise understanding of the mechanism of action as well as potential causes of side effects. It also enables more intelligent subsequent screening of compounds with better, more relevant assay readouts," he says.

Iconix Pharmaceuticals in Mountain View, in

collaboration with MDS Pharma Services of Lincoln, Nebraska, applies chemogenomics to the study of comprehensive responses of rats to treatment by a compound. Iconix uses CodeLink microarrays from Amersham Bioscience in Piscataway, New Jersey, which carry 10,000 probes for rat genes and expressed sequence tags, to measure gene expression in tissues from rats that have been treated with different drugs.

MDS and Iconix have identified more than 250 clusters of genes whose expression changes in response to 600 drugs. "These clusters of genes, or 'drug signatures', describe and predict many specific mechanisms of action of the drug in the whole animal without the caveats of *in vitro* systems," says Pauline Gee, vice-president of predictive biology at MDS. Iconix's DrugMatrix database now includes more than 200 of these drug signatures, which are used by drug companies such as Bristol-Myers Squibb to help select candidate drug molecules more efficiently and to improve their understanding of the safety of new compounds before preclinical development.

C.S.



Pathway visualization in Iconix's DrugMatrix.

especially as it relates to disease processes, to be explored in an efficient manner.” Using a mouse model of a trait associated with a human disease, his group identifies regions of the mouse chromosome that control susceptibility to the disease-associated trait. Gene-expression profiling with the mouse microarrays is also done on target tissues from the mouse strains under study.

“The differentially expressed genes encoded within the identified chromosomal regions are the candidate genetic susceptibility loci,” says Peltz. “We then analyse the function of the candidate genes (and pathways) in the biology related to the disease.” A mouse model of osteoporosis was recently developed using this method, and an enzyme identified that affects bone mass and bone quality, which led to the development of a new potential therapeutic for treating osteoporosis.

IntegraGen’s GenomeHIP technology platform takes a different microarray-based approach to discovering genes linked to complex disease traits in humans. The technology uses DNA chips for genomic comparisons of pairs of related individuals with the same disease to discover areas in which the genomes are identical — on the assumption that at least one of the relevant disease genes in such individuals will be identical in both. Crucial to IntegraGen’s technology is the prior removal of most of the polymorphisms that differentiate the two genomes, cutting down the amount of genome that has to be searched. The company claims that its method is cheaper and faster than conventional methods of gene identification, such as linkage analysis

using microsatellite markers. Unlike such methods, which can require several generations of many different families, the GenomeHIP technology can identify potential disease genes with high statistical significance in a relatively small sample of patients (50–150 pairs of related individuals with the same trait) in less than nine months.

Using this platform, IntegraGen recently identified a mutant G-protein-coupled receptor that is strongly associated with obesity in various populations. “The most frequent mutation observed was demonstrated to impair the signalling through this receptor by its natural ligands,” says Mous. “The normal function of this receptor, which is predominantly expressed in the gut, is to transmit an anorexic signal, which could be impaired in patients expressing the mutant form.”

The advent of the complete human-gene chips is generally welcomed as an important tool. But Fitzgerald cautions that, like most molecular tools, gene chips have limitations of which one must be aware. “Experimental design with an eye towards rigorous statistical analysis is crucial,” he says. “Gene chips produce a large number of individual gene patterns and if one is asking a computer program to find patterns in the data, it will find patterns. If the experiments were not



Jan Mous: faster mapping.

correctly designed and controlled, those patterns can be quite misleading.”

He also stresses the importance of remembering that in general, gene chips measure correlative rather than causative events. Andrea Gnirke of Xantos Biomedicine in Munich, Germany, also makes the point that the chips provide an analytical challenge. “Chip data can only be correlative data and implicate an enormous effort for data analysis, and it is difficult to sort out the ‘good’ targets that have causative effects in diseases,” she says.

And because much cellular regulation occurs at the level of proteins, “increased transcript production as measured by gene chips does not always correlate with production of protein, and even if more protein is produced it may not be active because it requires post-translational modification or relocalization within a cell”, cautions Fitzgerald. Jean-Jacques Yarmoff of Hybrigenics agrees: “Modifications of proteins lead to an additional level of complexity that is not necessarily addressable by human-gene chips.”

And because much cellular regulation occurs at the level of proteins, “increased transcript production as measured by gene chips does not always correlate with production of protein, and even if more protein is produced it may not be active because it requires post-translational modification or relocalization within a cell”, cautions Fitzgerald. Jean-Jacques Yarmoff of Hybrigenics agrees: “Modifications of proteins lead to an additional level of complexity that is not necessarily addressable by human-gene chips.”

Tackling the ‘interactome’

Because interactions between proteins are key to their function, determining the pathways within which a potential drug target

GETTING UP SPEED

Multiplexing in highly automated systems is one way of speeding up the assays used for target identification. Bio-Rad in Hercules, California, offers Bio-Plex, a multiplex analysis system that allows the simultaneous analysis of up to 100 different biomolecules (proteins, peptides or nucleic acids) in a single microplate well. Bio-Plex integrates assay kits, software, calibration and validation tools, and instrumentation into a complete system. The Bio-



Andrea Gnirke: high-throughput gene screening.

Plex suspension array and workstation uses three main technologies: a family of fluorescence-tagged microspheres from Luminex in Austin, Texas; a flow cytometer that uses two lasers to measure biochemical reactions occurring on the surface of the microspheres; and a high-speed digital signal processor to manage the fluorescent output.

In its target-identification programme, Xantos Biomedicine of Munich, Germany, uses high-throughput cellular screening assays to study gene function in parallel for thousands of genes. For whole-genome screening, its proprietary XantoPrep and XantoScreen high-throughput system screens complementary DNA libraries by overexpression in transient transfection systems, enabling Xantos to carry out functional analysis on up to 100,000 different cDNAs per month. “Genes or proteins are selected by their ability to directly influence physiologically relevant processes or pathways in cellular assays, leading to a very small number of relevant hits compared with expression-profiling studies, for example,” says Andrea Gnirke, senior scientist for expression analysis at Xantos. These methods have been used to identify secreted proteins influencing the differentiation of fat cells, which play a major role in the development of insulin resistance, and which Xantos hopes could have potential for the treatment of obesity-related type II diabetes. An interesting future direction for Xantos, along with many other companies, includes RNA interference (RNAi). “Xantos is implementing a confocal high-content imaging screening system and aims to integrate RNAi technology for high-throughput application,” says Gnirke.

C.S.

acts in the cell and the interactions it makes with other proteins are crucial. To learn more about these complex interactions, researchers are starting to map and study the 'interactome', the comprehensive listing of which protein interacts with which in an organism. A recent breakthrough by Curagen researchers is the production of an interactome for *Drosophila melanogaster*, the first comprehensive protein-interaction map for a multicellular organism. They used a high-throughput version of the yeast two-hybrid method to identify more than 20,000 unique interactions involving 7,000 genes in *Drosophila*. Interactions within and between protein complexes were also studied using mathematical modelling.

The study points to a connection between proteins known to be altered in human diseases and the 'druggable' classes of enzymes—those enzymes known to bind and be altered by small-molecule drugs. This is a relief for those trying to find potential therapies for human diseases such as cancer, heart disease and diabetes. Many of the interactions discovered in the *Drosophila* study will now be examined further in human cells for their relevance to disease. Curagen has obtained a similar set of human protein–protein interaction data by the same methods, but is keeping them under wraps for now.

In another approach to target discovery, Curagen has developed a group of targets it calls the 'pharmaceutically tractable genome' (PTG). This is a collection of about 8,000 genes whose products are predicted to make good targets as judged by one of several criteria: they are present in blood or on cell surfaces, rendering them accessible to drugs and antibodies; or are in a class of intracellular proteins that can be modulated by small molecules. This approach allows researchers to focus resources on investigating targets that are both novel and yet likely to lead to successful drugs.

Paris-based Hybrigenics also uses functional genomics for target identification, based on a proprietary yeast two-hybrid method that evaluates which proteins interact with the protein of interest, and whether the interaction is statistically significant. Yarmoff, senior director of business development, says that the company's approach also provides functional information. "Because we obtain many fragments which interact with the yeast two-hybrid bait, we can compute the intersection of these fragments to define the selected interacting domain (SID)," he says. "This is extremely useful information, as by comparing the SID with known functional domain sequences, we can understand not only which protein our bait interacts with, but also the function of the domain at which this interaction takes place. The richness of this information is what allows us to go beyond simple target



Aviva's SealChip16 for patch-clamping (above) is the heart of its PatchXpress7000 system (left) for assaying ion channels.

gives micro-positioning of cells and tight high-gigohm seals.

identification to use this information for target validation." The large amounts of data generated by this approach can be visualized using Hybrigenics' PIMRider viewer, which enables these protein-interaction networks to be explored.

Channel hopping

Ion channels are the basis of neural function and so are attractive drug targets but, historically, screening for drugs that affect ion channels has been a bottleneck in target identification because the tests cannot be automated successfully—the quality of the data obtained has not been good enough. The alternative, electrophysiological screening of individual cells by a researcher, produces high-quality data, but at a pace that is slow by drug-development standards, and is extremely monotonous and tedious to do. Advances in this area are much needed, because although analysis of the human genome has predicted that ion channels could account for up to a quarter of all potential new targets, only 5% of drugs currently on the market affect ion channels.

To address this problem, Aviva Biosciences in San Diego, California, has developed its SealChip16 system. This makes high-throughput, high-quality, patch-clamp recordings from single cells, in conjunction with Aviva's PatchXpress7000, a high-throughput patch-clamp workstation developed in collaboration with Axon Instruments of Foster City, California. Aviva's system incorporates a high-throughput planar patch-clamp technology that

Future challenges

"One of the major challenges in discovery is the complexity of the interactions that result in normal or disease states," says Seth Cohen of Caliper Life Sciences in Mountain View, California. "In the past, a simplistic view was, by necessity, taken, which resulted in many drugs failing in preclinical or clinical trials for lack of efficacy or side effects. The newer approach must account for this complexity. The holistic systems-biology approach to research will be necessary to overcome this challenge."

Mous agrees: "A challenging new development in the field of drug-target discovery is systems biology, or the recognition that genes, or better the gene products, are part of, and function, in large complex networks. Understanding the malfunctions in these large pathways will allow the selection of the best drug target for compensating the malfunction of the homeostasis of a biochemical process in a diseased person."

Peltz stresses that studying target pathways will be critical to identifying a target that is important in disease. "The major challenge is the lack of 'connectivity' with pathways for most of the genes in the genome," he says. Studying the disease pathway is the best way to find targets that are most suitable for drug development, he adds. Unfortunately, this is nearly impossible to do at present, because we do not know enough about most genes' interactions. This lack of information, says Peltz, "makes many of the genes we identify as associated with a disease 'isolated islands' that we cannot connect with the 'mainland'. Hopefully, the protein-interaction network maps that are being prepared in various organisms will help to fill this gap."

Caitlin Smith is a science writer based in Portland, Oregon.

Company	Products/activity	Location	URL
Proteomics			
Abgent	Polyclonal and monoclonal antibodies; customized antibody development; off-the-shelf and custom proteins and peptides; protein-expression services	San Diego, California	www.abgent.com
Active Motif	TransAM ELISA kits for transcription-factor assays; antibodies; peptide nucleic acids for gene silencing; kits and reagents for cell fractionation; nucleic acid isolation	Carlsbad, California	www.activemotif.com
Beyond Genomics	Target validation using a systems biology approach	Waltham, Massachusetts	www.beyondgenomics.com
Biacore	Analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
Ciphergen	ProteinChip system for protein identification and characterization from complex mixtures using SELDI technology	Fremont, California	www.ciphergen.com
Inpharmatica	Biopendium and CeleraEdition Biopendium proteome annotation resources; PharmaCarta	London, UK	www.inpharmatica.com
Hybrigenics	Target identification from protein interaction pathways and experimental validation services	Paris, France	www.hybrigenics.com
Jerini Array Technology	Substrate identification service for kinases using pepSTAR peptide microarray technology	Berlin, Germany	www.jerini.com
KeyNeurotek	Functionality and tissue-based screening for target identification, validation and compound/drug testing. Disease models available for glutamate-induced neurodegeneration and ischaemia/stroke	Magdeburg, Germany	www.keyneurotek.de
MDS Proteomics	Target validation using proteomics-based systems biology	Toronto, Canada	www.mdsproteomics.com
MorphoSys	Antibodies by Design custom monoclonal antibodies; HuCAL GOLD library	Martinsried, Germany	www.antibodyservices.com
PanVera	Protein reagents and assays for receptor and enzyme activity	Madison, Wisconsin	www.panvera.com
Pepscan	Chip with 1,200 protein kinase targets (peptides)	Lelystad, the Netherlands	www.pepscan.com
Phenomix	Identification of disease models, drug targets and validation for immune, metabolic, neurological and respiratory diseases	La Jolla, California	www.phenomixcorp.com
Proteome Factory	Target identification and validation	Berlin, Germany	www.proteome-factory.de
QED Bioscience	Catalogue and custom antibodies; immunoassay kits; 'DNA to antibody' custom monoclonal and polyclonal antibodies by mouse genetic immunization	San Diego, California	www.qedbio.com
TissueInformatics	TissueAnalytics high-throughput automated histology laboratory service, analytical software; drug target discovery	Pittsburgh, Pennsylvania	www.tissueinformatics.com
Upstate	KinaseProfiler specificity testing, IC50Profiler Express and services for crystallography assay development, bulk protein production and purification; chromatin immunoprecipitation kits; signalling proteins	Charlottesville, Virginia	www.upstate.com
Xerion Pharmaceuticals	XStream proprietary antibody-based target inhibition technology	Martinsried, Germany	www.xerion-pharma.com
Zyomyx	Human cytokine profiling biochip system	Hayward, California	www.zyomyx.com
Genomics and gene-expression profiling			
Aclara BioSciences	LabCard microfluidics 'lab-on-a-chip' systems for biochemical and cellular analysis, including multiplexed gene-expression analysis	Mountain View, California	www.aclara.com
Affymetrix	GeneChip microarray systems, including human whole-genome DNA microarray; microarray scanners and readers	Santa Clara, California	www.affymetrix.com
Agilent	Workstation for RNA, DNA or protein analysis; human whole-genome DNA microarray; instruments for genomics and proteomics	Palo Alto, California	www.agilent.com
Althea Technologies	PCR-based high-throughput gene-expression profiling	San Diego, California	www.altheatech.com
Capital Genomix	GeneSystem 320 differential gene-expression profiling; custom antibodies; microarray services	Chantilly, Virginia	www.capitalgenomix.com
Celera	Discovery System web-based informatics tools for accessing the Celera annotated genomes databases; bioinformatics services; DNA sequencing; target validation	Rockville, Maryland	www.celera.com
Chromagen	Fluorescence based products, kits and services for gene expression and protein detection and assay	San Diego, California	www.chromagen.com
Curagen	GeneScape portal for genome analysis tools; proteomics research for drug target discovery; proteome interaction map for <i>Drosophila</i>	Branford, Connecticut	www.curagen.com
GeneFormatics	Genomics-based drug target discovery	San Diego, California	www.geneformatics.com
Genomic Solutions	Automated instrumentation for genomics and proteomics, software, consumables and services. Investigator automated system for proteomics	Ann Arbor, Michigan	www.genomicsolutions.com
Hybridon	Cyclicons synthetic fluorescent DNA probes	Cambridge, Massachusetts	www.hybridon.com
Iconix Pharmaceuticals	DrugMatrix databases and software platform for chemogenomics-based research for drug discovery	Mountain View, California	www.iconixpharm.com
Immunosol	Ribozyme-based Inverse Genomics platform for identification of drug target genes	San Diego, California	www.immunosol.com
Incyte Genomics	Annotated gene and expressed sequence tag databases; Proteome BioKnowledge Library species-specific protein information databases; bioinformatics software	Palo Alto, California	www.incyte.com
IntegraGen	GenomeHIP (Genome Hybrid Identity Profiling) gene-mapping technology; target discovery	Evry, France	www.integragen.com
NimbleGen Systems	Custom and off-the-shelf microarrays; human whole-genome DNA microarray	Madison, Wisconsin	www.nimblegen.com
RNA-TEC	Custom oligonucleotide synthesis	Leuven, Belgium	www.rna-tec.com
Sigma-Genosys	Tools for molecular biology; oligonucleotide libraries for gene expression profiling for human, mouse, rat, <i>Bacillus subtilis</i> , zebrafish, <i>Escherichia coli</i>	The Woodlands, Texas	www.sigma-genosys.com
Xantos Biomedicine	Disease-specific and other cDNA libraries	Munich, Germany	www.xantos.de
Transgenics			
Ingenium Pharmaceuticals	Custom transgenic mouse and rat models; drug discovery	Martinsried, Germany	www.ingenium-ag.com
Lexicon Genetics	Gene-knockout and gene-function databases; transgenic mice for target identification and validation	The Woodlands, Texas	www.lexgen.com
memorec Stoffel	Custom transgenic and knockout mice for target identification and validation; cDNA microarrays; SAGE analysis	Cologne, Germany	www.memorec.com

Company	Products/activity	Location	URL
Nucleis	Custom transgenic mice	Lyon, France	www.nucleis.com
OzGene	Custom transgenic and knockout mice for target identification and validation	Canning Vale, Australia	www.ozgene.com
Zygogen	Transgenic zebrafish system for target identification and validation	Atlanta, Georgia	www.zygogen.com
Gene knockdown and RNA interference			
Ambion	Silencer range of kits, vectors, siRNAs and reagents for RNAi; kits and products for RNA synthesis, isolation, quantitation and analysis	Austin, Texas	www.ambion.com
atugen	siRNA-based drug discovery; custom phosphorothioate antisense RNAs and siRNAs for drug target identification and validation	Berlin, Germany	www.atugen.com
Benitec	Proprietary gene silencing RNAi technology	St Lucia, Queensland	www.benitec.com.au
Cenix BioScience	Genome-based high-throughput applications of RNA interference; siRNA design	Dresden, Germany	www.cenix-bioscience.com
Dharmacon	siRNA kits and arrays; custom RNAs; large-scale RNA synthesis	Lafayette, Colorado	www.dharmacon.com
Eurogentec	Custom siRNA for RNAi	Seraing, Belgium	www.eurogentec.be
GeneTools	Morpholino antisense oligonucleotides	Philomath, Oregon	www.gene-tools.com
Imagenex	Plasmid-based GeneSuppressor siRNA kits for RNA interference	San Diego, California	www.imgenex.com
InvivoGen	siRNA cloning vectors; custom siRNA in plasmid and viral vectors	San Diego, California	www.invivogen.com
Mirus	Products for gene transfer and RNAi	Madison, Wisconsin	www.genetransfer.com
MWVG Biotech	Custom oligonucleotide synthesis; Dharmacon siRNA kits and arrays; custom sequencing; microarrays and oligo sets	Ebersberg, Germany	www.mwv-biotech.com
New England Biolabs	HiScribe RNAi transcription kit	Beverly, Massachusetts	www.neb.com
Novagen	RiboJuice siRNA transfection reagent	Madison, Wisconsin	www.novagen.com
OligoEngine	pSUPER RNAi vector system for gene silencing; custom oligonucleotides for all applications	Seattle, Washington	www.oligoengine.com
Promega	Vectors, reagents and kits for genomics, proteomics and cellular analysis; kits for siRNA production and delivery	Madison, Wisconsin	www.promega.com
Proligo	Custom and speciality DNA and RNA oligos; siRNA oligos for RNAi	Hamburg, Germany	www.proligo.com
QIAGEN	Custom siRNAs and transfection reagents; Cancer siRNA Oligo Set Version 1.0	Venlo, the Netherlands	www.qiagen.com
Sequitur	Antisense and siRNA compounds	Natick, Massachusetts	www.sequiturinc.com
Spring Bioscience	Knockdown kits for producing siRNAs from target DNA sequence; products for gene cloning, target screening and protein-expression analysis	Fremont, California	www.springbio.com
Xeragon	siRNA design tool; custom siRNA synthesis	Germantown, Maryland	www.xeragon.com
General			
Amersham Biosciences	Instruments, equipment and products for genomics, proteomics and cellular assays; workstations and reagents for DNA microarrays; IN Cell Analyser confocal imaging system for rapid cellular assays	Piscataway, New Jersey	www.amershambiosciences.com
Apogent Technologies	Labware and equipment for the life sciences	Portsmouth, New Hampshire	www.apogent.com
Applied Biosystems	DNA sequencers, protein sequencers, mass spectrometers, chromatography systems, PCR systems, peptide synthesizers, nucleic acid synthesizers	Foster City, California	home.appliedbiosystems.com
Aviva Biosciences	SEAL Microchips for high-throughput patch-clamp ion-channel measurements	San Diego, California	www.avivabio.com
Axon Instruments	Scanners and microarray software; ImageExpress for live-cell assays	Foster City, California	www.axon.com
BD Biosciences	Culture media, radioassay kits, FACS range of flow cytometers; monoclonal antibodies, antibody arrays for proteomics; reagents and biochemicals for molecular biology	Franklin Lakes, New Jersey	www.bd.com
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics	Fullerton, California	www.beckmancoulter.com
Bio-Rad	Products, instruments and software for life-sciences research; Bio-Plex system for multiplex assays	Hercules, California	www.bio-rad.com
Caliper Technologies	LabChip 'lab-on-a-chip' microfluidic systems for high-throughput screening for drug discovery, biological and genetic research	Mountain View, California	www.calipertech.com
Cellomics	ArrayScan HCS and KineticScan HCS systems for automated cell-based high-content screening for drug discovery	Pittsburgh, Pennsylvania	www.cellomics.com
Chemicon	Off-the shelf and custom antibodies; reagents and kits for biochemistry and molecular biology	Temecula, California	www.chemicon.com
Cytomx	Antibodies, cDNAs and cell lines for molecular and cell biology; services for genomics and proteomics, cell-line production	Cambridge, UK	www.cytomx.com
Fluidigm	Microfluidics chip for protein crystallization	San Francisco, California	www.fluidigm.com
Gyros	CD microlaboratory for sample preparation for protein mass spectrometry	Uppsala, Sweden	www.gyros.com
Hamilton Company	Microlab automated liquid-handling workstations	Reno, Nevada	hamiltoncomp.com
LumiCyte	Protein biomarker profiles for drug development	Fremont, California	www.lumicyte.com
Luminex Corporation	xMAP platform for microbead-based multiplex assays	Austin, Texas	www.luminexcorp.com
Invitrogen	Kits and reagents for genomics, proteomics, molecular and cell biology; Lipofectamine 2000 for siRNA transfection	Carlsbad, California	www.invitrogen.com
MDS Pharma	Services company for all aspects of drug discovery	Lincoln, Nebraska	www.mdsp.com
PerkinElmer Life Sciences	Automated systems for liquid handling and sample preparation; high-throughput human DNA microarrays; HydroGel BioChip	Boston, Massachusetts	lifesciences.perkinelmer.com
Proteodyne	BioCube automated systems for research and genome-based drug discovery	Windsor, Connecticut	www.proteodyne.com
Roche Diagnostics	Reagents and kits for molecular biology, functional genomics and proteomics research	Lewes, UK	www.roche-applied-science.com
Tecan	Robotic laboratory workstations and automated solutions for genomics and proteomics	Medford, Massachusetts	www.tecan.com
Stratagene	Tools and reagents for molecular biology, genomics, proteomics, drug discovery and toxicology; GeneEraser siRNA transfection reagents	La Jolla, California	www.stratagene.com

● see advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejp.com

naturejobs

The waiting game

Young scientists at Georgetown University in Washington DC heard a tale of woe last month that had them nodding in recognition. Talking to a group of neuroscience postgraduates, Stacie Grossman, an associate editor at *Nature Medicine*, recalled her time in the lab at Georgetown. She had spent a long time crafting a grant proposal and, when it was finished, it seemed to her to be “beautiful”. She was so proud — and protective — of her submission that she drove to the US National Institutes of Health (NIH) in Bethesda, Maryland, to hand deliver it. She presented the proposal to the appropriate programme officer. “And they threw it onto this pile of 6,000,” Grossman said.

But that was the emotionally easy part. Waiting for an answer was painful, she said. The wait, and the accompanying uncertainty, was one of the reasons she began a career in scientific publishing.

The postgraduates at the meeting agreed that waiting — whether it be for a manuscript to be accepted at a journal or for a grant from the NIH — is one of the worst aspects of being a young scientist. “There’s this constant fear of rejection,” said graduate student Jill Weisberg. “Then there’s the actual rejection.”

The students said that handling known unpleasantness — such as low stipend levels or long lab hours — is easy compared with the uncertainties of funding, manuscript acceptance or whether they will ever be able to get onto the tenure track. “A lot of people lose focus because their grant is so delayed,” said Gregory Einch.

Postgrads basically have two choices. Accept and embrace uncertainty, or look for a scientific career that seems more solid. But scientific institutions can help, too, by being quicker with their responses. Because a rapid negative answer may, for some, be less traumatic than a drawn-out positive one.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Reinventing the Silk Road p236

CAREER VIEW

Bricks & Mortar
Building Copenhagen’s
BioCentre
Graduate Journal
The lab environment
Movers
Enric Banda

p238

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Can the intellectual route from China to the United States become a two-way street? Paul Smaglik investigates.

Modern genetics reached China in 1937, when Jiazhen (“C. C.”) Tan went home with a PhD from Caltech. It fell out of favour during the 1950s, but Tan lived to see his Institute of Genetics at Fudan University restored to prominence in the 1970s. In 1948, Edinburgh-trained physiologist Robert Kho-Seng Lim — head of the Chinese Red Cross during the Second World War — left China for the United States, where he was to become the first US National Academy of Sciences member of Chinese descent.

Most of the traffic was one-way at first, as scientists fled to the West during the civil war of the 1940s or the Cultural Revolution (1966–76). China has traditionally been indifferent to the outside world; there was little cultural exchange with western Europe, and the United States rarely allowed its citizens to visit China after 1949, when the Communists came to power. That changed in the late 1970s, when China began promoting scientific exchange in an effort to modernize and diplomatic relations were restored with the United States.

The ‘science road’ has become more of a two-way street recently. Although many Chinese scientists who find success abroad decide to stay there, a high percentage visit home regularly. In the past few years, increasing numbers have either moved back to China or visited for longer stints to teach courses, establish new institutes and modernize existing ones.

FOREIGN EXCHANGE

Ray Wu, a biochemist at Cornell University in New York, helped establish in 1982 the China–United States Biochemistry and Molecular Biology Examination and Administration programme (CUSBEA), which brought hundreds of top scientists to the United States. About 90% stayed in the United States, says Wu, but they now support visiting scientists who will go back and help China ‘catch up’.

“I feel that I have been privileged with the chance to study and work in this country,” says Wu, who



arrived from China in 1949 and has returned there twice a year since 1980.

China has enormous potential: in the United States alone, Chinese students have earned 40,000–50,000 PhDs in the past 25 years, adds Wu.

“If everyone goes back to China there’s no reservoir to hold all this talent,” says molecular biologist Xiang-Dong Fu, who considers himself a member of the “first class of the Cultural Revolution”, having been among the first to enter the reopened universities shortly after it ended. He was in the first group to go to a US university under Wu’s CUSBEA programme, earning his PhD at Case Western University, then doing a postdoc at Harvard under Tom Maniatis. Like many of his contemporaries, Fu intended to go back until the Tiananmen Square political clamp-down happened in June 1989.



K. SU/CORBIS

Reinventing the Silk Road (above): Ken Chien (far left) and Chung-I Wu (left) are hopeful that improvements to China's scientific facilities will help to attract expertise back to the country.

"Because of political turmoil and instability, we got stuck here," Fu says. The US Congress issued what many Chinese Americans of his generation call the 'June 4 Green Card', which gave permanent residence to all Chinese scholars in the United States.

Fu says the tide is turning in China, with schemes such as the Yangtze River Scholars paying young scientists well above the market rate. But the country needs to attract and retain more senior scientists — ones who are effectively able to write grants, conduct meetings and publish in international journals. "Most of those people are still in the United States," Fu says.

A few larger universities are succeeding. Tan's Fudan University recently recruited Li Jin, from the University of Cincinnati, Ohio, as dean of life sciences, and he has put the school up for an independent international review. Meanwhile, it is bolstering its genetics team by bringing in Yale *Drosophila* researcher Tian Xu and two researchers from the University of Colorado, Boulder: Min Han, who works on *Caenorhabditis elegans*, and mouse geneticist Yun Zhuang. The Chinese Academy of Science scored a coup when it attracted Mu-Ming Poo from Berkeley to run its Institute of Neuroscience.

Western-trained scientists have also taken turns to lecture 300 graduate students at a time in Shanghai and Beijing universities.

Yi Rao — now an associate professor of anatomy and neurobiology at Washington University School of Medicine — came to the United States in the 1980s, with the help of Charles Dickens rather than CUSBEA. "I read a few issues of *Nature*, *Science* and *Cell* and figured I couldn't do good science in China," says Rao, who pored over classic nineteenth-century fiction to improve his English.

Three years ago he initiated a course in molecular cell biology in Shanghai, taught in tandem with his China-based partner Jiarui Wu. They have now expanded the course to Peking University and its neighbouring rival Tsinghua University.

Such courses could help Chinese scientists in both the East and the West, says Rao. Chinese students are eager to pick up new techniques, but they do not always know how best to apply them.

TAIWAN INDICATOR

Chung-I Wu, chairman of the Department of Ecology and Evolution at the University of Chicago, agrees that Chinese students need more encouragement to think creatively, but adds that they have plenty of energy and ambition. He hopes that China's growing tendency to conduct large-scale science and engineering — such as building massive dams and conducting high-profile space launches — doesn't eclipse smaller, more elegant projects. "We have a joke about it," says Wu. "If you cannot make something beautiful, you make it big."

Wu sees parallels between Taiwan, where he is from, and mainland China. He visits the two countries four or five times a year. Taiwan is about 10–15 years ahead in terms of repatriation, he says. But he feels it is not the best model as Taiwan has a top-down management structure, which China is emulating, and an emphasis on seniority, which, so far, China is avoiding.

His compatriot Ken Chien, now director of the Institute of Molecular Medicine at the University of California, San Diego, agrees that Taiwan is seeing a stream of scientific repatriation, compared with China's trickle. But this could change, because facilities in China are improving. "They've put an emphasis on a few top universities," says Chien. "I think that's good."

China is using the market economy to draw back top scholars, sometimes offering US-level salaries. "If you have a US salary and you're living in China, you are doing very well," says Chien.

In mainland China, says Chien, administrators do not worry about paying a young scientist "ridiculously more" than senior scientists in the same department — the benevolent side of the top-down system.

Changes in Chinese scientific culture are not always for the better. Within the past ten years, there has been a shift to publish in top international journals. The downside is what many see as an excessive dependence on citations indexes.

What is still missing, say Chien and Ray Wu, is a rigorous peer-review system. Two years ago, Wu helped to set up the National Natural Science Foundation of China — a sort of US-style peer-review body — which is helping to improve the grant system. In Taiwan, Chien set up a similar process ten years ago.

Modernization may take time, but it is happening. "I have learned to have patience," Yi Rao says. ■

Paul Smaglik is editor of *Naturejobs*.

GRADUATE JOURNAL

The lab environment

My current lab is a great place to work. The people here are very proficient in their fields, and everyone is friendly, open and helpful — which means that I enjoy the time during (and after) work. People were also very open at a lab where I worked some years ago, but I didn't enjoy its competitive environment and the responsibility that imposed.

I believe that a working environment that matches my character makes a big difference to my performance in the long run. I think that when you are considering starting a project in a lab you have not been to before, you should imagine spending a much longer time there. How would you feel after a few years there?

But how can you get a feel for whether you would enjoy working somewhere? One simple strategy is to talk to as many people as you can when you visit for an interview. And maybe you have a friend of a friend who knows someone who works there who will be able to give you an inside view.

The working climate and culture are just as important if you are choosing a job in industry. I actually consider it to be as important as the salary, even though most people give this the most attention after spending time in academia. ■

Philipp Angerer is a second-year PhD student in biotechnology at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

BRICKS & MORTAR

Building Copenhagen's BioCentre

Work will begin this year on the Copenhagen BioCentre, which will cost 640 million kroner (US\$105 million) and be funded by the Danish government. This facility, which is due to open its doors in 2006, will provide lab and office space for 500 employees from three research institutions focusing on molecular biology and biomedical research.

One of these three institutions is the recently established Biotech Research and Innovation Centre (BRIC). The Danish Ministry of Science, Technology and Innovation established BRIC to form an international centre of excellence in molecular biology and biotechnology, strengthening the research environment in the Scandinavian Medicon Valley area (see *Naturejobs*



Claus Bræstrup (left) and Kristian Helin bring research and industry together.

10–11; 21 June 2001). Until it moves to the Copenhagen BioCentre in 2006, BRIC is based at the Symbion science park in Copenhagen.

BRIC is a consortium of leading Danish research institutions and is associated with the University of Copenhagen. Its aim is to establish collaborations between public research institutions and industry, and to promote the exchange of ideas within the Danish biotech research community. To achieve this end, it is managed by an independent board

of directors from academic institutions and industry.

BRIC's first research group, established in 2003, focuses on developing functional screens to identify novel genes involved in cancer, and on using mouse genetics to understand protein function and the development of cancer.

BRIC is still growing, and will have some 30 employees by the end of this year and about 200 when it moves into the Copenhagen BioCentre in 2006. ■

Kristian Helin is director of BRIC and Claus Bræstrup is its chairman.

MOVERS **Enric Banda, director-general, Catalanian Foundation for Research, Barcelona, Spain**



Like many scientists-turned-administrators, Enric Banda became a manager almost by accident. He gave up a well-funded research post in Switzerland early in his career, partly because he felt homesick for his native Spain. The price he paid for the ticket home was that he had to build a research programme in Catalonia from scratch. Over the years, he has learned to relish taking on such challenges.

That's not to say that these

challenges don't come without frustrations, Banda says. As director of the European Science Foundation in Strasbourg, France, for the past five years, he has learned to cultivate patience, as changes at the European level can only be made "at a speed that it is difficult for me to swallow", he says.

But even that had its rewards, says Banda. He learned that the reason for the lack of alacrity was that each European country has its own individual approach to science and culture. "You realize the diversity we have and how rich we are in approaches," he says. "But sometimes that makes life difficult."

Banda says that he grew to enjoy management — once he was resigned to the time it takes to bring about changes. "As a scientist, I used to say 'this is politics. I'm not interested'," he recalls. But he adds that it is gratifying to be able to give colleagues extra resources.

By the time his fixed term at the European Science Foundation ran out, he had had what he considers to be some success in finding funding for projects in some of the member countries. He also made the institution slightly faster at making decisions. "But it could be faster still," Banda says.

His new post as director-general of the Catalanian Foundation for Research in Barcelona, Spain, is just as challenging, but for different reasons. In a situation reminiscent of his earlier return to Spain, Banda must once again build a new body from the ground up. He would like to turn the Catalanian Foundation for Research into a think-tank for science policy and use it to create links between the public and private sector.

But he is glad to be home again, even though he enjoyed his spells in Zurich and Strasbourg. "I'm still a Mediterranean," Banda says. ■

CV **1998–03:** Secretary-general, European Science Foundation, Strasbourg
1988–91: Director, Institute of Earth Sciences, Spanish Council for Scientific Research (CSIC), Barcelona
1987–: Research professor, CSIC
1983–87: Head of geophysics, Geological Survey of Catalonia, Barcelona
1980–83: Research scientist, Institute of Geophysics, Swiss Federal Institute of Technology (ETH), Zurich